Journal of Applied Psychology

KENNETH E. CLARK, Editor
University of Colorado

Table of Contents

Personal History Correlates of Physical Scientists' Career Aspirations: Lewis E. Albright and J. R. Glennon..... 281

Simple Formula Aids for Understanding the Joint Action of Two Predictors: Melvin R. Marks, Raymond E. Christal, and Robert A. Bottenberg..... 285

A Factor Analysis of Job Activities: George J. Palmer, Jr. and Ernest J. McCormick..... 289

Use of Analytical Information concerning Task Requirements to Increase the Effectiveness of Skill Training: James F. Parker, Jr. and Edwin A. Fleishman..... 295

Operator Performance on a Chord Keyboard: H. C. Ratz and D. K. Ritchie..... 303

Noise, the "Arousal Hypothesis," and Monotonous Work: William N. McBain..... 309

A Verification Scale for the Strong Vocational Interest Blank, Men's Form: Robert W. Filbeck and Robert Callis..... 318

Personality and Product Use: W. T. Tucker and John J. Painter..... 325

Factor Analytic Definitions of Vocational Motivation: John O. Crites..... 330

Individual and Group Correlates of Attitudes toward Work-Related Change: Don A. Trumbo 338

Reliability, Sex Differences, and Validity in the Leaderless Group Discussion Technique: Walter A. Kaess, Sam L. Witryol, and Richard E. Nolan..... 345

American Psychological Association

Volume 45, Number 5 October 1961

Consulting Editors

ARTHUR BRAYFIELD, *Pennsylvania State University*
GEORGE E. BRIGGS, *Ohio State University*
NORMAN FREDERIKSEN, *Educational Testing Service*
LEONARD D. GOODSTEIN, *University of Iowa*
EDWIN R. HENRY, *Standard Oil Company of New Jersey*
JOHN HOLLAND, *National Merit Scholarship Corporation*
CLIFFORD E. JURGENSEN, *Minneapolis Gas Company*
LAURENCE S. MCGAUGHRAN, *University of Houston*
QUINN MCNEMAR, *Stanford University*
HAROLD F. ROTHE, *Fairbanks, Morse and Company*
THOMAS A. RYAN, *Cornell University*
A. C. WELCH, *University of New Mexico*

This journal gives primary consideration to original investigations in any field of applied psychology except clinical psychology, although a descriptive or theoretical article may be accepted if it represents a special contribution in an applied field. Quantitative investigations of interest or value to psychologists working in the following broad fields will be considered: vocational and educational prognosis, diagnosis, and guidance at the secondary and college level; personnel research in business, industry, and government; engineering psychology; industrial working conditions; research on opinion and morale factors; job analysis and classification research; market and advertising research.

Because of the large number of manuscripts submitted, authors should adhere to the rule of "brevity consistent with clarity." The typical manuscript should run to approximately 4,000 words. There is a lag of approximately twelve months between receipt and publication of an article. Authors may request advanced publication if they are prepared to pay the cost of printing the necessary extra pages.

Manuscripts should be addressed to the Editor:

Dr. Kenneth E. Clark
Office of the Dean
College of Arts and Sciences
University of Colorado
Boulder, Colorado

All manuscripts must be submitted in duplicate. Original figures are prepared for publication; duplicate figures may be photographs or pencil-drawn copies.

Manuscripts must conform to the style requirements described in the *Publication Manual of the American Psychological Association*.

Journal of Applied Psychology

Published bimonthly by the
American Psychological Association
Prince and Lemon Sts., Lancaster, Pa.
and 1333 Sixteenth Street N.W.
Washington 6, D. C.

$10.00 per volume $2.00 per issue

ARTHUR C. HOFFMAN
Production Manager

HELEN ORR
Promotion Manager

Subscriptions, orders, and business communications should be addressed to the American Psychological Association, 1333 Sixteenth St. N.W., Washington 6, D. C. Address changes must reach the subscription office by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

Second class postage paid at Lancaster, Pennsylvania and at additional mailing places.

© 1961 by the American Psychological Association, Inc.

Journal of Applied Psychology

VOL. 45, NO. 5 OCTOBER 1961

PERSONAL HISTORY CORRELATES OF PHYSICAL SCIENTISTS' CAREER ASPIRATIONS

LEWIS E. ALBRIGHT AND J. R. GLENNON[1]

American Oil Company

In discussing the problems of managing a research and development laboratory, several writers (Bello, 1956; Miles & Vail, 1960; Torpey, 1960) have mentioned the trend toward providing two ladders of advancement for the physical scientist employed in large industrial laboratories. One of these lines of progression is the supervisory or administrative route within the laboratory. The other consists of "bench" research positions of increasing complexity or responsibility, but requiring little or no administrative supervision of others. This latter hierarchy is intended to provide comparable "rewards" to productive researchers without diverting their efforts into administration. The rewards are, of course, larger salaries, more rank, impressive titles, and privileges similar to supervisory and managerial "extras."

Implicit in this modified organizational structure is the suggestion of basic differences, at least in career values and aspirations, between research scientists and administrative personnel. Common educational backgrounds, identical treatment during recruitment, and initial laboratory employment would tend to mask such differences, however, and these could be critical in subsequent performance. If differences do exist, then selection and placement techniques should be capable of development to assist in matching scientists and laboratory career choice.

One technique which has shown promise in a similar area—engineering job placement—is the Strong Vocational Interest Blank. Specifically, Dunnette (1957) found that scoring keys derived from the Strong discriminated among engineers working in different functions (pure research, sales, etc.). More recently, Kulberg and Owens (1960) showed that a number of life history items were differentially correlated with Dunnette's keys, which is not surprising since many "life history" items have to do with interests. But this finding does suggest that a personal history questionnaire itself might prove to be a valid discriminator of supervisory- vs. research-oriented scientists. The present study is an attempt to investigate that possibility.

METHOD

A detailed description of the personal history questionnaire, subjects, and procedures of data collection and analysis used in this study may be found in an earlier report (Smith, Albright, Glennon, & Owens, 1961). Briefly, a 484-item instrument covering various background topics was administered to employed petroleum research scientists; item responses were analyzed and cross-validated against three criteria of job success (supervisory ratings made separately on both over-all job performance and creativity, as well as the objective criterion of number of patent disclosures filed during a 5-year period).

For the purpose of the present study, the subjects were grouped according to their responses to the first two alternatives of Question 179 of the personal history questionnaire, which follows:

179. Assuming a free choice which line of progression would you prefer?

1. Research assignments, such as, research assistant, research associate, etc.
2. Supervisory or administrative assignments, such as, group leader, section leader, etc.
3. Advancement outside the Research Department
4. Any of the above

[1] The writers wish to acknowledge the assistance of Wallace J. Smith of the Standard Oil Company (Indiana) and William A. Owens of Purdue University in the conduct of this study.

Initially, only those individuals were considered who had not yet achieved positions high enough to be in either of the lines of advancement described in Responses 1 and 2 above. Therefore, these subjects were termed research and administrative *aspirants*, respectively. There were 79 of the former and 62 of the latter. The groups were equated for length of company service by eliminating 11 of the longer service research aspirants, resulting in the total complement of both groups having hiring dates after World War II. The groups, as then constituted, were compared with respect to their amount of education, ratings on job performance and creativity, and number of patent disclosures.

Following these comparisons, each group was subdivided into random halves and item analyses of the remaining 483 personal history items performed on each combination of research and administrative subsamples separately. Items were retained which discriminated the administrative from the research aspirants in *both* item analyses at the .10 level or beyond; on a compound probability basis this is equivalent to approximately the .05 level or less.

Results

Comparability of aspirant groups on education and criterion variables. Inspection of data on the educational attainment of the research and administrative aspirants revealed no essential difference between the groups on this variable. Inasmuch as the subjects were volunteers from the finite population of the laboratory, the t test proposed by Johnson (1949, p. 71) was used to test the significance of the difference between means of the groups' performance and creativity ratings, as given in Table 1. The results show that the administrative aspirants were rated significantly higher than the research aspirants on both over-all performance and creativity.[2] A median test (Walker & Lev, 1953) revealed that the median difference of six patent disclosures in favor of the research aspirants, also shown in Table 1, was significant. (The variation in Ns appearing in Table 1 is due to missing criterion data.)

Personal history analysis. The item analyses of the personal history questionnaire against the research-administrative dichotomy established from Item 179 revealed a total of 43 items which discriminated the aspirant groups at the required significance level. For illustrative purposes, three of these are given below, the "A" and "R" designating the group, administrative or research, making the indicated response more frequently than the other group.

133. How much time, on the average, do you spend reading technical or professional journals, magazines or books?

1. None
2. Less than 1 hour a week
A. 3. 1 to 3 hours a week
4. 4 to 7 hours a week
R. 5. More than 7 hours a week

150. Would your choice of an ideal job be one which:

A. 1. Allowed a great amount of interaction with other people
2. Would require working with a small group
R. 3. Would allow you to work closely with one other person
R. 4. Would allow you to work by yourself

436. Assuming you had sufficient musical ability and training to perform in the following ca-

[2] Because the performance ratings were based on a seven-step forced distribution system with 1 as the highest rating and 7 the lowest, the larger mean of 4.47 for research aspirants is in fact a poorer rating.

TABLE 1

Comparisons of Research and Administrative Aspirants on Performance and Creativity Ratings and Patent Disclosure Records

Variable	Research Aspirants			Administrative Aspirants			Significance		
	N	M	SD	N	M	SD	Test	df	p
Over-all Performance Rating	68	4.47	1.46	62	4.03	1.28	$t = 2.12^b$	128	$<.05$
Creativity Rating	58	−2.91	85.96	55	24.85	78.37	$t = 2.00^b$	111	$<.05$
No. of Patent Disclosures	42	7.00[a]	15.11	35	1.00[a]	10.02	$\chi^2 = 7.53$	1	$<.01$

[a] These values are medians, others are means.
[b] t tests for significance of difference between means from a small finite population (Johnson, 1949, p. 71).

pacities, which *one* do you believe would give you the greatest personal satisfaction:

1. Soloist—instrumental or vocal
R. 2. Composer
A. 3. Conductor
4. Arranger
5. Member of orchestra or choral group—not soloist
6. Critic

Although many of the significant items are more concerned with factual biographical information than are these three, there is a surprising confirmation of the stereotype of the research scientist as a somewhat introverted, work-oriented person vs. the more outgoing, status seeking, administrator.

Validity Generalizations

It is necessary, for prediction purposes, to verify these findings on independent samples because of the "internal" method of item validation used. Consequently, the records of 65 individuals not involved in the previous analyses were selected for study. Fifty-three of these subjects were either group leaders or section leaders, the remaining 12 being research associates or their assistants. Reference to Question 179 shows that their job titles placed these subjects in one of the two lines of advancement mentioned in Responses 1 and 2 and hence, are called administrative and research *incumbents,* respectively, in the remainder of this paper. Any incumbent not indicating a preference, on Question 179, for the line of advancement in which his job title placed him was held out for separate analysis.

Examination of the three criterion variables for the incumbent groups showed that the trends noted earlier for the aspirants remained the same, i.e., administrators were rated higher on performance and creativity but had fewer patent disclosures than the researchers. (Significance tests were not employed in this case because of limited data on one or more criteria.)

The validity of the 43 previously significant personal history items was investigated for the incumbents by use of a simple scoring scheme. Each incumbent, regardless of actual job title, was arbitrarily given a score of + 1 for every response characteristic of the *research* aspirants and a − 1 for an *administrative* type response. Table 2 shows the distribution of total scores on this scoring key. The means of 6.42 for researchers and − 1.72 for administrators are in the expected direction and differ significantly ($t = 5.32$ with 63 df, $p < .01$).

TABLE 2

Scores of Research and Administrative Incumbents on 43-Item Personal History Key

Total Score	Research Incumbents ($N = 12$)	Administrative Incumbents ($N = 53$)
15–16	X	
13–14		
11–12	X	
9–10	XX	
7–8	XXX	X
5–6	X	XX
3–4	X	XXXX
1–2	X	XXXXXXXXX
−1–0	XX	XXXXXXXXXXX
−3–(−)2		XXXXXXXXX
−5–(−)4		XXXXXX
−7–(−)6		XXXXXX
−9–(−)8		XX
−11–(−)10		X
−13–(−)12		X
M	6.42	−1.72
SD	4.70	4.05

An additional bit of evidence of the validity of the key was obtained by applying the scoring system to the papers of six administrative incumbents whose responses to Question 179 indicated they preferred the research line of advancement. Their mean score of 3.33, roughly midway between those of the other two groups, makes sense intuitively since these individuals see themselves as misplaced on their present jobs.

A second validity generalization study was carried out with individuals one or more levels higher in the organization than those utilized heretofore. Specifically, the personal history questionnaires were scored of five senior research associates and 12 division directors. These jobs are virtually the highest ones in the research and administrative hierarchies, respectively. (Obviously, the pyramidal shape of the organization seriously limits the number of such positions.) The responses of these groups were available on only 18 of the 43 discriminating items, but the means of .80 for the researchers and − 1.50 for the administrators follow the pattern previously estab-

lished. The small numbers of subjects preclude the meaningful application of any statistical test of significance; it would seem, however, that in confirming the trend, this finding is more than merely suggestive.

Discussion

The process of applying a scoring key developed at one level of the organization to successively higher echelons for validity generalization purposes is somewhat novel in the writers' experience. It seems reasonable to believe, however, that differences in personal history which can be discriminated at the entry level should persist as people move up, although such differences will be increasingly difficult to demonstrate due to smaller *N*s, greater remoteness of early experience for the older person, etc. How fruitful the method would prove in closely allied measurement areas such as interest and personality test development is questionable. With relatively homogeneous groups it would seem to offer real promise. But among managers in general, for example, the work of Porter and Ghiselli (1957), using a self-description adjective checklist, suggests that differences between management levels might mask within-level differences.

Besides their importance in job placement, significant personal history differences between industrial researchers and their administrator colleagues could have implications for other features of the laboratory climate. For example, it is the administrators who, by nature of their jobs, determine the performance ratings, salary adjustments, promotional prospects, and the like. They or their subordinates do much of the college recruiting. In short, biases toward favoring one's own kind, consciously or unconsciously practiced, could do a great deal of harm to the organizational careers of research scientists and, of course, serious disservice to the research achievements of the laboratory.

Summary

A total of 43 personal history items were found to differentiate employed petroleum research scientists desiring to advance in the laboratory supervisory hierarchy from a similar group aspiring to increased salary and status while remaining "at the bench." For purposes of validity generalization a scoring key composed of the discriminating items was applied successively to groups of scientists working at the middle and upper levels in the administrative and technical lines of advancement. The key retained its discriminating power (though to a lesser extent) at each higher echelon of the laboratory organization. Some implications of this result for personnel placement and management of an industrial research and development function were discussed.

REFERENCES

Bello, F. Industrial research: Geniuses now welcome. *Fortune*, 1956, **53**(1), 96–99, 142–150.

Dunnette, M. D. Vocational interest differences among engineers employed in different functions. *J. appl. Psychol.*, 1957, **41**, 273–278.

Johnson, P. O. *Statistical methods in research*. New York: Prentice-Hall, 1949.

Kulberg, G. E., & Owens, W. A. Some life history antecedents of engineering interests. *J. educ. Psychol.*, 1960, **51**, 26–31.

Miles, S. B., Jr., & Vail, T. E. Thinking ahead: Dual management. *Harvard bus. Rev.*, 1960, **38**(1), 27, 30, 149–150, 152–154.

Porter, L. W., & Ghiselli, E. E. The self perceptions of top and middle management personnel. *Personnel Psychol.*, 1957, **10**, 397–406.

Smith, W. J., Albright, L. E., Glennon, J. R., & Owens, W. A. The prediction of research competence and creativity from personal history. *J. appl. Psychol.*, 1961, **45**, 59–62.

Torpey, W. G. Conserving our technological manpower. *Personnel*, 1960, **37**(2), 61–67.

Walker, Helen M., & Lev, J. *Statistical inference*. New York: Holt, 1953.

(Received December 14, 1960)

Journal of Applied Psychology
1961, Vol. 45, No. 5, 285–288

SIMPLE FORMULA AIDS FOR UNDERSTANDING THE JOINT ACTION OF TWO PREDICTORS[1]

MELVIN R. MARKS

Psychological Research Associates, Arlington, Virginia

RAYMOND E. CHRISTAL, AND ROBERT A. BOTTENBERG

Personnel Laboratory, Wright Air Development Division

Suppose you had the problem of predicting academic success in a college course. You try two predictors in a large sample and find that one has a validity of .10 and the other has a validity of .00. Furthermore, you determine that these two predictors correlate .99 with each other. What would you conclude?

If your conclusion is that you should throw away the two predictors and start over again, then we suggest you read this article in its entirety. If, on the other hand, you are not surprised to learn that these two predictors will produce a multiple correlation of .707, then you might wish to skip to the formula derivations in the last section of this article.

It all started when, in actual research, a multiple R of .544 for two predictors was computed. Now there is nothing unusual about a multiple correlation validity coefficient of .544, except in this case the validity coefficients for the two predictors were .060 and .168, respectively. At first, the results were questioned—especially when it was noted that the correlation between the two predictors was .979. The investigator was familiar with so-called "suppressor effects," but he had not had occasion to see them in action to this extent.[2] He found himself checking the results on a calculator. A mistake had not been made.

When the finding was discussed with associates, their reactions were interesting. The statisticians took the attitude of "I told you so," for they had been warning researchers against forecasting multiple correlations from inspection of the predictor matrix. The attitude of most investigators, on the other hand, was one of "Well, I'll be damned." They realized they would not have predicted these results from inspection of the data. In fact, if they had come across two experimental variables with such a high intercorrelation, they would have discarded one without hesitation—especially if neither predictor had much relationship with the criterion.

We thought the reader might enjoy, and perhaps profit from some experiences the authors had while studying the effect of the intercorrelation term on the two-predictor, multiple correlation coefficient. Before continuing, let us adopt the following notation:

V_1 = validity coefficient of Predictor 1
V_2 = validity coefficient of Predictor 2
r = correlation between Predictor 1 and Predictor 2
R = multiple correlation

The first and primary point is that for any fixed pair of validity coefficients, there must be a value for the intercorrelation term which will produce a minimum, multiple correlation coefficient. Applying a little calculus to the basic two-predictor formula, the authors were pleasantly surprised to find that this value turns out to be simply the ratio of the two validity coefficients. That is:

$$\frac{V_1}{V_2} \text{ or } \frac{V_2}{V_1}$$

= value of the intercorrelation term for a fixed pair of validity coefficients in the two-predictor equation which results in a minimum R. Select the ratio which results in a value of 1.00 or less.

[1] The research reported in this paper was sponsored in part by Personnel Laboratory, Wright Air Development Division, under ARDC, Project 7734.

[2] The authors have some doubts about the term and normal explanation of "suppressor effects." The factorial explanation breaks down rapidly as the number of predictors increases beyond two, and is not completely adequate when there are only two predictors. For example, in the problem "from actual research" mentioned in this paper, V_1 is an aptitude composite and V_2 is the square of that same composite. It is difficult to explain what is being suppressed out of what in this instance.

Stated in other words, for the three-variable case, when the correlation between predictors is identical to the ratio of the predictor validity coefficients (smallest coefficient in the numerator), the multiple R is a minimum. Any value of the intercorrelation different from this ratio will yield a larger multiple R.

What is the value of this minimum R? With a little thought it becomes obvious that the minimum R always will be exactly equal to the larger of the two validity coefficients. In other words, the ratio of the two validity coefficients will locate for us an intercorrelation value such that the variable having the lowest validity will add nothing to the variable having the highest validity in predicting the criterion. Thus, when the intercorrelation is equal to the ratio of the validity coefficients, there is no increase in predictive efficiency by employing the variable with the lowest validity.

Since the ratio V_1/V_2 seems to be all-important for the effect of the intercorrelation on the magnitude of multiple R, it will be of interest to examine the results for the cases where $V_1 = V_2$, where $V_1 \neq V_2$ (neither V_1 nor V_2 is zero), and where V_1 or V_2 (but not both) is equal to zero. Note that the first and third cases are special, in that they rarely occur.

In the first case, where $V_1 = V_2$, the ratio V_1/V_2 is exactly 1.00. If two validity coefficients are equal, then, the smallest multiple correlation will be obtained when the two predictors are perfectly related. Any other value for the intercorrelation term will result in a higher multiple R.

"So what?" one might ask. Most of us have been taught that, in general, the lower the relationship between two predictors, the higher the resulting multiple. Perhaps the best answer is provided by a look at Case 3, where V_1 or V_2 (but not both) is equal to zero. In this instance V_1/V_2 is .00, and the lowest multiple R will be computed when the two predictors correlate zero. Thus, the mirror image of Case 1 exists, and it can be shown (see Table 1) that the higher the correlation of the predictors, the higher will be the resulting multiple R.

Both Case 1 and Case 3 are special cases. The more usual situation is Case 2, where $V_1 \neq V_2$ and neither is zero. For example, let us assume that V_1 is .50 and V_2 is .10. In this instance, V_2/V_1 is .20. If the correlation of Predictor 1 with Predictor 2 were .20, then Predictor 2 would add nothing to Predictor 1, and the resulting multiple R would be exactly .50. If the correlation of Predictor 1 with Predictor 2 were different from .20, in either direction, the resulting R would be greater than .50.

This simple formula could be especially helpful to investigators who lack ready access to computing facilities. Suppose, for example, one had a matrix containing the intercorrelations and validities for an operational selection composite and certain experimental variables. He could simply take the ratio of each of the experimental variables, in turn, to the validity coefficient of the selection composite. By comparing these ratios to the observed intercorrelation he could determine if any one of the experimental variables could add predictive variance to the selection composite. The greater the discrepancy between the ratio and the intercorrelation, the greater would be the expected contribution of the experimental variable to the composite. It should also be noted that situations may exist where one might make greater progress toward improving prediction by increasing the correlation between predictors (validities being held constant) rather than decreasing such correlations.

A formula to determine the values for the correlation between two predictors which will lead to a maximum R turns out to be very simple:

$$V_1V_2 \pm \sqrt{(1 - V_1^2)(1 - V_2^2)}$$

= two values for the correlation between two predictors (with fixed validity coefficients) which will grant multiple R of 1.00.

From applying this formula, we see that, when two predictors have validity coefficients of .50 and .10, respectively, they will produce a multiple correlation of 1.00 if their correlation with each other is either .91 or $-.81$.

A little more understanding of the influence of the correlation between two predictors on the resulting R can be obtained by inspecting Table 1. This table shows the fallacy of the widely accepted opinion that in order to get a high multiple correlation we should have a

TABLE 1

EFFECT OF THE CORRELATION (r) BETWEEN TWO PREDICTORS ON THE MAGNITUDE OF THE MULTIPLE CORRELATION COEFFICIENT FOR FIXED VALIDITIES V_1 AND V_2

r	Fixed Pair of Validity Coefficients						
$V_1 =$ $V_2 =$	.00 .50	.10 .50	.50 .50	.00 .80	.24 .80	.36 .60	.21 .30
−.86	.98						.96
−.84	.92						.90
−.82	.87						.85
−.80	.83	.97					.81
⋮							
−.70	.70	.80					.66
−.60	.62	.71		1.00			.57
−.50	.58	.64	1.00	.92		.97	.51
−.40	.55	.60	.91	.87		.89	.47
−.30	.52	.56	.85	.84	.95	.82	.43
−.20	.51	.54	.79	.82	.90	.77	.41
−.10	.50	.52	.75	.80	.86	.73	.38
.00	.50	.51	.71	.80	.84	.70	.37
.10	.50	.50	.67	.80	.82	.67	.35
.20	.51	.50	.64	.82	.80	.64	.34
.30	.52	.50	.62	.84	.80	.63	.33
.40	.55	.51	.60	.87	.80	.61	.32
.50	.58	.53	.58	.92	.82	.60	.31
.60	.62	.56	.56	1.00	.85	.60	.30
.70	.70	.61	.54		.92	.61	.30
⋮							
.80	.83	.71	.53			.63	.30
.82	.87	.74	.52			.64	.31
.84	.92	.77	.52			.65	.31
.86	.98	.82	.52			.67	.31
.88		.87	.52			.69	.32
.90		.95	.51			.73	.33
.92			.51			.77	.34
.94			.51			.85	.37
.96			.51			.98	.41
.98			.50				.52
.99			.50				.71
r for minimum R	.000	.200	1.000	.000	.300	.600	.700
r #1 for $R = 1.00$	−.866	−.812	−.500	−.600	−.390	−.530	−.870
r #2 for $R = 1.00$	.866	.912	—	.600	.774	.962	.996

high correlation between each of the independent variables and the dependent variable, and a low correlation between the predictor variables. That this viewpoint is not necessarily correct was demonstrated very clearly by Horst (undated). Although Table 1 illustrates the manner in which the R changes as the predictor intercorrelation varies for a few selected pairs of validities, and the formulas presented above provide a ready answer to the question of the values of r which minimize and maximize R in a two-predictor case, the reader should recognize that the only sure way of determining the extent to which a predictor will increase the predictive efficiency of a set of predictors is to compute the multiple correlation both with and without the predictor in question.

Formula Derivations

1. To solve for the value of r which minimizes R for a fixed pair, V_1 and V_2, differentiate the following expression with respect to r; then set the derivative equal to zero and solve the resulting equation for r.

$$R^2 = \frac{V_1^2 + V_2^2 - 2V_1V_2r}{1 - r^2}$$

$$\frac{dR^2}{dr} = \frac{(1 - r^2)(-2V_1V_2) - (V_1^2 + V_2^2 - 2V_1V_2r)(-2r)}{(1 - r^2)^2}$$

$$= \frac{-2V_1V_2r^2 + 2(V_1^2 + V_2^2)r - 2V_1V_2}{(1 - r^2)^2}$$

Setting the derivative equal to zero and multiplying through by $-(1 - r^2)^2$ we obtain:

$$2V_1V_2r^2 + 2(V_1^2 + V_2^2)r + 2V_1V_2 = 0$$

Solving for the roots of the equation, we obtain:

$$r = \frac{2(V_1^2 + V_2^2) \pm \sqrt{4(V_1^2V_2^2)^2 - 16(V_1^2V_2^2)}}{4V_1V_2}$$

$$= \frac{2(V_1^2 + V_2^2) \pm 2(V_1^2 - V_2^2)}{4V_1V_2}$$

$$r^2 = V_1/V_2, \qquad V_2/V_1 \qquad [1]$$

2. To obtain values of r which provide a multiple R of 1.00 when two predictors with fixed validity coefficients are least-squares weighted, solve the following equation for r:

$$\frac{V_1^2 + V_2^2 - 2V_1V_2r}{1 - r^2} = 1$$

This may be expressed as:

$$r^2 - 2V_1V_2r + V_1^2 + V_2^2 - 1 = 0$$

therefore, solving for r:

$$r = \frac{2V_1V_2 \pm \sqrt{4V_1^2V_2^2 - 4(V_1^2 + V_2^2 - 1)}}{2}$$

and simplifying:

$$r = V_1V_2 \pm \sqrt{(1 - V_1^2)(1 - V_2^2)} \qquad [2]$$

Restriction 1: When $V_1 = V_2$, only one root of Formula 2 is applicable and it reduces to the expression $2V^2 - 1$.

Restriction 2: Formula 2 is not applicable when $V_1 = V_2 = 0$. This is the only fixed pair of validity coefficients for which no value of r will grant a multiple R of 1.00.

Summary

It was demonstrated that the ratio of any two fixed validity coefficients will locate a value for the intercorrelation term in the two-predictor, multiple correlation formula such that the variable having the lower validity will contribute nothing to the variable having the higher validity. The resulting multiple correlation is thereby a minimum.

A second formula is presented that determines two intercorrelation values for any fixed pair of validity coefficients which will result in a multiple correlation of 1.00. This second formula does not hold when both validity coefficients are exactly zero, and is restricted to determination of a single intercorrelation value when the two validity coefficients are exactly equal and not zero.

REFERENCE

Horst, P. The role of prediction variables which are independent of the criterion. In Social Science Research Council, Committee on Social Adjustment, *Prediction of personal adjustment*. New York: SSRC, undated. Pp. 431–436.

(Early publication received March 9, 1961)

Journal of Applied Psychology
1961, Vol. 45, No. 5, 289–294

A FACTOR ANALYSIS OF JOB ACTIVITIES[1]

GEORGE J. PALMER, JR. AND ERNEST J. MCCORMICK

Louisiana State University *Purdue University*

Over a period of years considerable progress has been made in the identification of the nature of different human abilities and characteristics. In the area of mental abilities, for example, the classic work of Thurstone (1947) is, of course, generally considered a milestone. In the domain of motor skills, some of the studies by Fleishman (1960) have helped to crystallize the nature of these human dimensions.

While some human attributes have thus become somewhat clarified, there has been relatively little effort given to the identification of what might be thought of as the dimensions of jobs. Viewing human work and its many variations, it is reasonable to believe that there might be certain basic dimensions of work activity which conceivably could be identified and described. A few studies have tended to confirm this hypothesis. Among such studies are those by Coombs and Satter (1949), Hemphill (1959), Besnard (1954), and McCormick, Finn, and Scheips (1957).

In approaching the domain of human work, however, one could take at least two different points of view. First, one could consider the description of work primarily in terms of job characteristics. Such descriptions might emphasize the technological aspects of jobs, or they might emphasize what workers *accomplish* in their jobs (such as baking, selling, painting, and so forth). Descriptions in terms of technology or in terms of accomplishments might be thought of as "job oriented" descriptions of the work activities of people. On the other hand, one could view human work in terms of what workers *do* in performing their jobs, e.g., visual, manual, or communications activities. These might be thought of as "worker oriented" descriptions of what the worker *does*, i.e., how he uses the technological factors to achieve certain ends.

[1] This study was supported by funds from the Purdue Research Foundation.

The distinction between the "job oriented" and "worker oriented" approaches is not always clear cut; it is largely a matter of emphasis. In general, perhaps the differences can be summarized by saying that the "job oriented" view is likely to place more emphasis on the conditions and results of work, while the "worker oriented" view tends to place more emphasis on worker activities as such.

This study was one that was viewed as a "preliminary" and "probing" effort in this direction. In particular, it was the plan to develop a check list of job activities of a "worker oriented" nature, to use these in describing a sample of jobs, and to subject the results to factor analysis to attempt to identify the worker activity "dimensions" of the jobs in question. In this exploratory effort it was realized that the experimental check list of work activities that might be developed would be only something of a "first approximation" of what might ultimately become a more precise device. It was further realized that, with what would be a restricted sample of jobs, the resulting "dimensions" would have to be viewed in the light of the sample of jobs in question and, therefore, would not necessarily have general applicability. In other words, this was viewed primarily as a methodological study to find out if it might be profitable to explore this relatively virgin territory.

PROCEDURE

Job Activities Check List

A 177-item check list, describing job activities in terms of worker behaviors, was organized generally along lines suggested by information theory notions, such as input, decision, storage, and output. The major parts of the check list and brief description of their contents are presented below. Most of the examples given are the identifying stems from items that were more fully defined and illustrated in the check list proper.

1. Information-Receiving Activities (35 items relative to: use of the senses to obtain information to

carry out work activities—for example, identifies or distinguishes sounds by pitch or tone)

2. Mental Activities (43 items relative to: decision making activities; uses of various kinds of knowledge, including general and specific vocational educational development—for example, evaluates performance of people)

3. Supervisory and Communications Activities (33 items relative to: responsibility for supervision; oral, written, and signal communications; interpersonal contacts involved in these activities—for example, supervises work group)

4. Manual Activities (25 items relative to: use of hands to operate controls, vehicles, tools, or equipment; direct use of hands to modify materials—for example, operates keyboard device, i.e., typewriter)

5. General Bodily Activities (24 items relating to: use of the whole body in simple or coordinated movements—for example, climbs)

6. General Work Conditions (7 items relating to: conditions in the physical work environment—for example, noise level)

7. General Job Characteristics (10 items relating to: ratings of the job as a whole on general worker attributes—for example, adaptability to routine)

Sample of Jobs

A job sample was selected to represent the job structure of a large steel producing firm. Of an estimated 10,000 jobs in the company, job descriptions were available for about 5,000, from which the sample of 250 was selected.

The composition of the sample was as follows: Exempt salaried jobs (operating, administrative, technical, and related jobs), 40%, were described in terms of purpose, function, skills, knowledge, and responsibility. Nonexempt clerical jobs, 10%, were described in terms of tasks and time distribution. Hourly paid jobs (steel production and processing, maintenance, inspection, and related jobs), 50%, were described concisely in terms of functions, education, supervision, and other conditions, as given by the Cooperative Wage Bureau (1953) for steel industry jobs.

It should be noted that the objects of study were jobs, not personnel, nor positions (personnel-in-jobs). The jobs were not sampled in proportion to numbers of incumbents, so that the present results cannot be interpreted in terms of importance for employed personnel.

Analysis of Jobs

Data and Method. The basic information about the jobs in the sample was contained in written job descriptions. Each job in the sample was analyzed on the basis of the information in its descriptions, in terms of the item of the check list. One of the investigators analyzed each job by reading its description, indicating on the check list the presence or degree of each job activity which he considered to be pertinent to the job in question.

Reliability of Analysis. The reliability of analysis was estimated from the degree of agreement between competent analysts for a sample of 32 jobs. Eight graduate students in industrial psychology made independent analyses of 2 to 5 jobs apiece, and their analyses were compared with those of the investigator. The index of agreement for dichotomous items was the coefficient of overlap, Peters and Van Voorhis (1940), Formula 61. For continuously scored items, the index was the Pearson product-moment correlation.

Reliability data were based on 130 items from the total list. Some items were eliminated because they were seldom or never checked. Thirty-two items were eliminated because the overlap coefficient was indeterminate, a result of a zero denominator.

The median coefficient of agreement was .75 and only 7% were below .41, indicating a reasonably satisfactory degree of argument among the analysts in selecting the check list items which they considered to be pertinent to each job.

Methods of Data Analysis

The check list data for the sample jobs were correlated and subjected to factor analyses in two stages.

Initial Factor Analyses. The first factor analyses were based on the correlations of check list variables within each of five groups of variables. These groups consisted essentially of the items in Sections 1 through 5 of the check list. Items were scored dichotomously and were intercorrelated with an index developed specifically for check list data by Ben J. Winer, Purdue University.[2] Items which were checked less than 15 times were not included in the analysis, so that the final matrices for Sections 1 through 5 contained from 21 to 31 items. Each matrix was factored separately by the multiple group method as given by Thurstone (1947). The highest coefficients in the columns or rows of each group were used as communality estimates.

Factor Scores. For purposes of estimating the factors, simple integral weights ranging from -3 to $+3$ were assigned to items according to the sign and magnitude of their factor loadings. For each factor variable, each job's score was the algebraic sum of weighted items checked for it. It would have been possible to develop a complete multiple regression equation to estimate each factor, but as

[2] For a 2×2 table, a|b/c|d, where a is the $++$ cell and letters are cell frequencies, this index is

$$I = \sqrt{\frac{|a - 1/2(b + c)|}{a + b + c}}$$

This coefficient takes the sign of the difference inside the absolute value bars. It may be noted, too, that it ignores Cell d and gives equal weight to b and c.

Guilford (1954), for one, points out, the gains are small for the labor.

Final Factor Analysis. For purposes of integrating the factors obtained from parts of the check list, a sixth factor analysis was carried out. Factor scores which had been developed for the 24 factors resulting from the factor analyses of Sections 1 through 5 were intercorrelated (using Pearson product-moment correlations) together with 14 items not hitherto included for analysis. Some of these 14 items were surviving items from Sections 6 and 7 concerning General Work Conditions and General Job Characteristics; also included were several items relating to educational development which had been reserved for inclusion in this final analysis.

A 28×28 correlation matrix was factored by the method of principal components programed for the Datatron computer. The highest correlations were used as trial communalities in a first factoring and, upon identification of the chief factors, re-estimated communalities were used in a second analysis. The decision to stop factoring was based on the rate of reduction of the residuals and rotation to orthogonal simple structure was effected by the graphic single-plane method. Jobs were given scores on the rotated, final factors variables by the weighting procedure previously described.

Results [3]

Results of Initial Factor Analyses

The five initial factor analyses resulted in 14 separate oblique group "factors." These factors were typically bipolar, because to some extent job activities had this mutually exclusive character, and partly because of the way items were defined.

In interpreting the factors, it should be noted that the basic data were descriptive of jobs, and that the resulting factors were characterized by combinations of coexistent activities. The basic data are perhaps best considered to be occupational (or perhaps organizational) in nature, rather than psychological. The factors, then, consisted of those activities which tended to "hang together" in jobs. Thus, it might be that two or more distinctly different "types" of activity might be common to the same factor because of the fact that these activities typically occurred together in jobs. This point is emphasized because factors from such data would not be expected to have the same kind of "psychological" sense that would be true of relationships among measures of human test performance.

The interpretation of the factors is indicated by the descriptive titles which are intended to summarize the general pattern of factor loadings. The multiple group factors are listed below according to the section of the check list from which they were derived. To illustrate each factor, sample jobs which typify jobs with high and low scores (in that order) on the factor are given in parentheses.

Information-Receiving Activities.[4] Factor 1: Visual Information from Physical Objects vs. Displays (Scrap Inspector vs. Copywriter)

Factor 2: Visual vs. Tactual Information from Objects (Welder vs. Fireman)

Mental Activities.[5] Factor 3: Routine vs. Administrative Decision Making (Car Checker vs. Farm Manager)

Factor 4: Supervisory and Professional Decisions vs. Physical Activity (General Foreman vs. Operator, Tube Rolling)

Factor 5: Routine vs. Business Planning Decisions (Hydrostatic Tester vs. Supervisor of Commercial Markets)

Supervisory and Communications Activities.[6] Factor 6: Routine Verbal Communications vs. Signals or Managerial Contacts (Personnel Clerk vs. Medical Director)

Factor 7: Communications Given Verbally vs. Via Signals (Head Timekeeper vs. Ambulance Driver)

Factor 8: Communications from Signals vs. Personal Contacts (Craneman vs. Market Specialist)

Factor 9: Originating vs. Receiving Communications (Superintendent of Foundry vs. Clock Repairman)

[3] A seven-page table containing items of highest loading and sample jobs for the multiple group factors has been deposited with the American Documentation Institute. Order Document No. 6738 from ADI Auxiliary Publications Project. Photoduplication Service, Library of Congress; Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

[4] For information purposes, the correlation between the group factors was $r_{12} = .35$.

[5] The correlations among the group factors were $r_{34} = -.30$, $r_{35} = .52$, $r_{45} = -.30$.

[6] The correlations among the group factors were $r_{67} = .72$, $r_{68} = .27$, $r_{69} = -.55$ $r_{78} = -.12$, $r_{79} = .64$, $r_{89} = -.32$.

TABLE 1

FACTOR LOADINGS

Variable	Loading
Factor I: General Decision Making and Mental Activity	
Reasoning	.91
Language	.91
Decision Making vs. Physical Activity	.85
Originating vs. Receiving Communications	.82
Latitude of Job Activities	.79
Mathematics	.72
Communications Given Verbally vs. via Signals	.70

FACTOR SCORE LEVEL	SAMPLE JOB
High	Medical Director
Average	Salesman
Low	Bricklayer Apprentice

Factor II: Sedentary vs. Physical Work Activity	
Sedentary vs. Physical Activity	.52
Visual Information from Physical Objects vs. Displays	−.46
Routine vs. Business Planning Decisions	−.47
Manual Operations	−.52
General Physical Activities vs. Postural Restrictions	−.63

FACTOR SCORE LEVEL	SAMPLE JOB
High	Interviewer
Average	Guard
Low	Hooker, Crane

Factor III: Communications in Business Management vs. Information in Routine Physical Work	
Originating vs. Receiving Communications	.31
Communications from Signals vs. Personal Contacts	−.38
Routine Verbal Communications vs. Signals or Managerial Contacts	−.39
Routine vs. Administrative Decision Making	−.44
Routine vs. Business Planning Decisions	−.48
Visual Information from Physical Objects vs. Displays	−.54

FACTOR SCORE LEVEL	SAMPLE JOB
High	Superintendent
Average	Clerk Typist
Low	Mechanical Repairman

Factor IV: Knowledge of Tools vs. Mathematics	
Use of Craftsman's Tools vs. Other Hand Implements	.44
Use of Body of Knowledge	.36
Routine vs. Administrative Decision Making	−.34
General Physical Activities vs. Postural Restrictions	−.37
Mathematics	−.52

FACTOR SCORE LEVEL	SAMPLE JOB
High	Instructor, Carpenter Apprentice
Average	Mill Shearman
Low	Payroll Clerk

Manual Activities.[7] Factor 10: Manual Operations (Operator, Tube Rolling vs. Head Timekeeper)

Factor 11: Use of Craftsman's Tools vs. Other Hand Implements (Millwright vs. Research Technician)

General Bodily Activities.[8] Factor 12: General Physical Activities vs. Postural Restrictions (Helper, Axle-Maker vs. Safety Inspector)

Factor 13: Sedentary vs. Physical Activity (Supervisor vs. Operator, Oiling Machine)

Factor 14: Manual vs. General Bodily Activities (Time Study Man vs. Hooker, Crane)

Results of Final Factor Analysis

In the final factor analysis, a principal components analysis extracted four factors which may be considered higher-order (although not second-order) factors; all were based in part on factors derived from the initial factor analyses, and in part on the 14 new variables introduced in this final analysis. The bipolar nature of some of these final factors resulted from two conditions: first, from the positive and negative loadings of the variables, and, second, from the bipolar nature of the multiple group factors correlated with the final factors. Principal (orthogonally rotated) loading and sample jobs are given for each final factor, in descending order of importance (see Table 1).

Factor I accounts for 66% of the variance and is a general factor. Factors II, III, and IV account for 15 to 8% of the variance. Factor I indicated that the greatest differences between jobs lie in Decision Making and Mental Activities. Factor II indicates differences in physical activity. Factor III refers to various sources and methods of communications for originating, relaying, or receiving job information. Factor IV suggests knowledge or skill in the use of mathematics, perhaps relating to manual and computational routines. The mathematical aspect of Factor IV is, of course, independent of Factor I on which mathematics also has a high loading.

The emergence of Factor I, with its emphasis on decision making and mental activities as the most dominant factor, seems to be in keeping with the results of a number of job evaluation studies which have suggested that the most important performance influencing the evaluation of jobs are those associated with mental requirements, skill, experience, and related variables (Tiffin & McCormick, 1958). Designation of this factor in terms of decision making activities gives it an interpretation consistent with the views of organization theorists, such as Simon (1948), who hold that, on the vertical dimension of organizations, the main difference between jobs is related to the level of decision making.

Limitations

On the whole, the results show that a large number of job activities for a rather wide range of jobs can be organized meaningfully in terms of a small number of independent dimensions. However, the results should be interpreted with certain limitations in mind.

First, it should be noted that the check list of worker oriented activities was viewed as a "first approximation" toward the development of what might ultimately be a much more refined device. Further, the sample of jobs was clearly a very limited one, the sample consisting of jobs from only one industry. It should be pointed out that the job descriptions varied in length, detail, and format, depending upon the class of jobs in question and, also, that the check list was used with the job descriptions (the secondary source), rather than with direct observation of the jobs themselves.

Conclusions

This study was carried out as something of a "probing" project to test the feasibility of identifying what might be thought of as the "dimensions" of worker activities of jobs. Specifically, factor analyses of five sections of the worker activities check list produced 14 multiple-group factors. A subsequent factor analysis of the jobs in terms

[7] The correlation between the group factors was $r_{10,\ 11} = .00$.

[8] The correlations among the group factors were $r_{12,\ 13} = -.11$, $r_{12,\ 14} = .34$, $r_{13,\ 14} = .23$.

of their scores on these 14 factors and 14 other selected variables produced four more general factors, as follows: General Decision Making and Mental Activity, a general factor; Sedentary versus Physical Work Activity; Communications in Business Management versus Information in Routine Physical Work; and Knowledge of Tools versus Mathematics.

In examining the original 14 factors and the four final factors, a number of them seem to make a reasonable amount of "logical" sense. Certain of the factors, however, seem to be rather distinctly influenced by what apparently were fairly unique facets of some of the jobs in the particular sample in question.

It may be concluded that the results tend to support the view that work activities can be identified or measured and that the variety of human work activities may be organized with greater simplicity and economy in terms of a smaller number of relatively independent dimensions.

REFERENCES

BESNARD, G. G. Shred-outs of tasks performed by senior B-29 mechanics. *USAF Personnel Train. Res. Cent. tech. Rep.*, 1954, No. 54–4.

COOMBS, C. H., & SATTER, G. A. A factorial approach to job families. *Psychometrika*, 1949, **14**, 33–42.

COOPERATIVE WAGE BUREAU. *Job description and classification manual for hourly-rated production, maintenance, and nonconfidential clerical jobs.* Pittsburgh: CWB, 1953.

FLEISHMAN, E. A. The description and prediction of perceptual-motor skill learning. Presented at Symposium on Training Research and Its Implications for Education, University of Pittsburgh, February 1–3, 1960.

GUILFORD, J. P. *Psychometric methods.* New York: McGraw-Hill, 1954.

HEMPHILL, J. K. Dimensions of executive positions: A study of the basic characteristics of the positions of ninety-three business executives. *Educ. Test. Serv. res. Bull.*, 1959, No. 59–5.

MCCORMICK, E. J., FINN, R. H., & SCHEIPS, C. P. Patterns of job requirements. *J. appl. Psychol.*, 1957, **41**, 358–364.

PETERS, C. C., & VAN VOORHIS, W. R. *Statistical procedure and their mathematical bases.* New York: McGraw-Hill, 1940.

SIMON, H. A. *Administrative behavior.* New York: Macmillan, 1949.

THURSTONE, L. L. *Multiple factor analysis.* Chicago: Univer. Chicago Press, 1947.

TIFFIN, J., & MCCORMICK, E. J. *Industrial psychology.* Englewood Cliffs, N. J.: Prentice-Hall, 1958.

(Received August 31, 1960)

Journal of Applied Psychology
1961, Vol. 45, No. 5, 295–302

USE OF ANALYTICAL INFORMATION CONCERNING TASK REQUIREMENTS TO INCREASE THE EFFECTIVENESS OF SKILL TRAINING[1]

JAMES F. PARKER, JR. AND EDWIN A. FLEISHMAN

Psychological Research Associates *Yale University*

The design of training programs for complex skills customarily involves a combination of expert judgment and known principles of learning. Often these programs are highly successful. However, to any psychologist involved with practical training problems, the gap between learning principles developed in the laboratory and the application of these to training practice becomes painfully apparent (DuBois & Manning, 1957; Gagné & Bolles, 1959; McGehee, 1958). There are many reasons for this including a lack of knowledge of the manner in which principles of learning are related to particular characteristics of the tasks learned. Furthermore, laboratory studies of learning frequently are concerned with treatment variables which are apt to have only a small practical effect in training situations relative to the large effects of individual difference variables (e.g., abilities) (DuBois, 1960; Fleishman & Hempel, 1955). The problems of task dimensions and ability differences are not unrelated and can be handled in the same conceptual framework. For example, tasks can be described in terms of the abilities required to perform them (Fleishman, 1960).

The present report evaluates a training program in which use was made of analytical information about ability requirements of the task to be learned. Two types of analytical information were used. The first involved knowledge about the more general abilities (e.g., spatial orientation) which contribute to performance on the task in question. The second type of information concerns the relative contribution of the different specific task components (e.g., control stick or pedal movements) to over-all task proficiency. In this latter case one can ask if the specific ability to use the pedals at a given stage of training is highly related to learning rate and to final proficiency; if so, can this information be used effectively in training? In applying information about both kinds of task components (general abilities and specific intratask relationships), special attention was given to their contributions to individual differences at different stages of training on the task in question.

We will proceed in the following order: We will describe the task to be learned, how the analytical information was obtained, the training program developed, and the evaluation of the program.

The Training Task

Display-Control Characteristics

The task to be learned was a complex tracking task, described in detail in previous reports (Parker & Fleishman, 1959, 1960). A device was constructed so as to simulate roughly the display characteristics and control requirements of an interceptor aircraft. The subject envisioned himself to be flying the attack phase of an airborne radar intercept mission. His task was to keep a target dot in the center of an oscillograph display while at the same time trying to keep a pointer centered on an instrument dial, designated a "sideslip indicator." The controls used by the subject involved a standard aircraft control system of stick and rudder pedals. These were coupled in a manner similar to those of an actual aircraft. Thus, if the target dot on the scope was to the right the subject made appropriate control movements to steer the "craft" to the right. These movements would bring him "on target" and the target dot would return to the center

[1] This study was performed under Contract Nonr-2489(00) between Psychological Research Associates and the Office of Naval Research. A more detailed account can be found in the original report by the same title prepared for ONR.

of the scope display. As in an aircraft, all such turning movements required coordinated action of stick and rudder controls. Thus, application of right control pressure without the proper amount of right rudder produced a sideslip to the left and a consequent left deflection on the "sideslip indicator." This indicator provided the subject with additional information regarding the "degree of coordination" he was achieving in centering the dot. The target course was programed through an analog computer. The control stick was an acceleration control; that is, the extent of the subject's stick movement was directly related to the acceleration of dot displacement. The rudder pedals approximated a pure velocity control. The task was a very difficult one for subjects to master. Photographs of the device may be found elsewhere (Parker & Fleishman, 1959, 1960).

Performance Scores

The following five performance measures were obtained during each testing session. Four of these were error scores; one was a time-on-target score. The error scores represent the extent of target dot (or indicator needle) displacement summated through time.

Integrated absolute error score. This was recorded at the conclusion of every trial and was produced by summing algebraically the three absolute error part scores described in accordance with this relationship:

$$T = 1/2X + 1/2Y + Z \qquad [1]$$

where: T = Integrated absolute error score
X = Absolute azimuth error
Y = Absolute elevation error
Z = Sideslip error

Each of the following part scores was recorded every third trial. An automatic sequencing mechanism switched from one part score to the next at the conclusion of each trial.

Azimuth (yaw) part score. A summation of the total error occurring in the horizontal axis of the oscillograph display

Elevation (pitch) part score. A summation of the total error occurring in the vertical axis of the display

Sideslip error part score. A measure of the lack of coordination between stick and rudder movements

Time-on-target score. The time during the trial when the target dot and the sideslip indicator needle were both (simultaneously) within prescribed tolerance limits

The Analysis of Tracking Performance

We now turn to the analysis made of the component task requirements for this particular tracking task. The rationale and details of this analysis have been reported elsewhere (Parker & Fleishman, 1959, 1960). We will summarize this briefly and then describe the application of this analysis to the experimental training program.

Derivation of Ability Requirements

The approach here was an extension of a series of studies by Fleishman (1957, 1960), Fleishman and Fruchter (1960), Fleishman and Hempel (1954, 1955), and Parker and Fleishman (1959, 1960). These studies use the fact of individual differences to probe into the ability requirements of a task. These studies have demonstrated that in learning a complex task the particular abilities contributing to proficiency may change systematically at different stages of practice. This suggests the possibility that a more efficient method of training might take into account this changing pattern of abilities.

Information regarding the changing ability requirements of the tracking task described above was obtained in a previous study (Parker & Fleishman, 1959). A comprehensive battery of 44 reference tests of known factorial structure was administered to 203 subjects who also mastered the tracking task. There were 17 tracking sessions, each consisting of 21 one-minute trials. The intercorrelations among the reference tests were factor analyzed. The tracking performance scores (integrated absolute error measure) at 10 selected stages of practice next were projected onto the rotated axes defined by the reference battery, using Mosier's extension method (Fruchter, 1954). This provided the factor loadings of each tracking

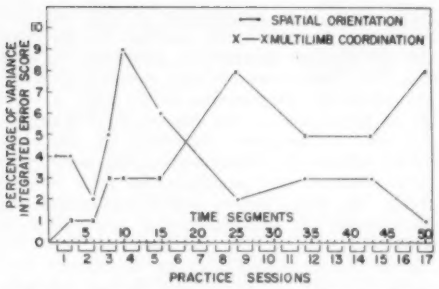

FIG. 1. Changes occurring in the importance of two ability factors with increased practice in tracking.

practice stage, and indicated the amount of variance in tracking at that stage of practice which could be accounted for by each ability factor.

Although 15 factors[2] were identified in the reference battery, only two, Spatial Orientation and Multilimb Coordination, contributed significant variance of a systematic nature to tracking performance.

Figure 1 presents the contribution of these two factors through the training program. It can be seen that after Practice Session 3,[3] there is a decrease in the contribution of Spatial Orientation to tracking proficiency. On the other hand, the Multilimb Coordination factor increases in importance up to session nine. We will return to the use made of this information in the subsequent experimental training program.

Determining the Contribution of Subtask Components

As described above, measures of component performances (azimuth error, elevation error, and sideslip error) were obtained along with measures of over-all performance (integrated error and time-on-target) during the course of practice. In a complex task such components are related to and form a part of over-all proficiency, although these precise relationships are seldom known. For the present task the correlations among all these component scores with each other and with the over-all proficiency measures were obtained for 10 different stages of practice (Parker & Fleishman, 1959). Of particular interest for the present study are the correlations of each component score, recorded at each stage of practice, with final over-all proficiency recorded in the last stage of practice (fiftieth time segment[3]). These correlations appear in Table 1.

It can be seen that how well a person does in the component performance (using stick, rudder, etc.) in the early stages of learning is not very predictive of final over-all proficiency. (It might be mentioned that early *over-all* proficiency did not predict final level either.) However, as practice continued prediction from all components increased, with the best predictions obtained from the "sideslip" score during later practice sessions. Since this component score reflected the extent of rudder pedal and stick coordination we have further confirmation of the factor analysis results, which indicated an increase in the contribution of the Multilimb Coordination factor.

On the assumption that knowledge of the relative contribution of component performances can provide guidance for training, this information was used in the experimental training program to be described next.

TABLE 1

CORRELATIONS OF COMPONENT SCORES AT VARIOUS STAGES OF PRACTICE WITH TERMINAL INTEGRATED ERROR SCORE

	Time Segment								
Error Score	1	3	6	8	10	15	25	34	43
Azimuth	02	−07	03	11	19	27	28	34	49
Elevation	08	10	12	17	26	34	36	42	52
Sideslip	12	04	−04	00	13	30	34	51	65

Note.—Decimal points omitted.

[2] The factors identified were: I. Spatial Orientation, II. Control Precision, III. Speed of Arm Movement, IV. Manual Dexterity, V. Reaction Time, VI. Verbal Comprehension, VII. Response Orientation, VIII. Arm-Hand Steadiness, IX. Perceptual Speed, X. Visualization, XI. Integration, XII. Pursuit Confusion Doublet, XIII. Mechanical Experience, XIV. Finger Dexterity, and XV. Multilimb Coordination.

[3] The 17 practice sessions were subdivided into 51 time segments in order to provide better definition of the learning curves. The abscissa of Figure 1 shows the relation of practice session to time segment and to the 10 time segments selected for use in the data analyses.

The Experimental Training Study

Three groups of subjects were compared. Groups I and II are regarded as control groups and Group III is the experimental group. These groups differed only in the nature of the verbal instructions administered during the course of practice.

All groups were drawn from the same population: freshmen Air Force ROTC students at the University of Maryland. All students were volunteers and were paid for their participation in the testing program. The training period was identical for both control groups and the experimental group: 17 sessions, distributed over 6 weeks, with each session consisting of 21 one-minute trials in tracking.

Group I—no formal training. The first control group ($N = 203$) was the group described previously. It provided the data for the correlational and factor analyses of tracking performance. This group received no formal training on the tracking task other than a brief explanation of the controls, the display, and the task. During the entire learning period these subjects received no instruction of any kind other than the answering of questions. They were told their score following each trial, however, in order that they might evaluate their individual progress.

Group II—"common sense" training. The second group ($N = 61$) was administered a "common sense" type of training program constructed to be as much like operational military training for this type of task as possible. Discussions were held with qualified navy personnel to insure that this type of training was realistic in terms of present navy practices. This group provided a basis for determining whether the experimental training represented an improvement over current practices when both were compared with Group I. This program involved an initial explanation and demonstration, guidance, and assistance to subject during early trials, and later individual practice with critiques following certain sessions.

Group III—the experimental training. It will be recalled that this study was based upon two premises:

1. When it is known that an ability is important at one point in the practice schedule, the verbal emphasis of that ability at an earlier point in time will facilitate learning.
2. A knowledge of the learning of component activities within the tracking task and the interrelation of these scores with terminal proficiency can be used as a means of structuring training.

The above premises were the basis of the experimental training program for Group III ($N = 60$).

As Figure 1 indicated, Spatial Orientation achieved maximum importance at the fourth tracking session. Therefore, training instructions were developed to be used prior to and during the third session of tracking. The specific instructions were developed in an attempt to give insight into the Spatial Orientation requirements of this task. The relationship between target position on the scope and the required direction of movement for both stick and rudder controls was described carefully. The Multilimb Coordination factor starts a rapid rise in importance at the end of the fifth session and rises to maximum importance at the ninth session (Figure 1). Therefore, training instructions were developed, for administration prior to and during the fifth session, which emphasized the simultaneous use of control stick and rudder in controlling movement of the target dot.

The component tracking score information presented in Table 1 was utilized in the following way for Group III. Prior to and during both the ninth and eleventh sessions (Time Segments 25 and 31, respectively [8]), instructions reemphasized the importance of centering the sideslip indicator and suggested the importance of monitoring that indicator rapidly and frequently while continuing to control the target dot on the oscillograph.

It should be noted that the experimental program administered to Group III was "overlaid" on the "common sense" training program which was administered to Group II. That is, all procedures used with Group II, such as initial indoctrination, early monitorship and guidance, and later critiques, were administered to Group III in conjunction with the experimental program. This was done in order to allow a direct evaluation of any improvement in training effectiveness as a result of the addition of the experimental portions.

Every effort was made to keep motivation high in all three groups. There was every indication that subjects in all three groups were highly motivated to do their best.

Results

A check on the initial matching of the three groups indicated the groups matched on time-on-target (see Figure 3), azimuth error, and elevation error measures. However, the integrated error and sideslip error scores indicated a slight initial superiority for Group III ($p < .05$). A check on this revealed this was due to an inadvertent instruction given these subjects during their initial session to center the rudder pedals. Since this was such an easy operation, it was felt that this effect was transitory and did not affect the results. This assumption seems tenable in view of the parallel results obtained with the time-on-target score (Figure 3), on which the groups were initially matched.

Figures 2 and 3 present learning curves for all three groups in terms of total "integrated error" and "time-on-target" measures.

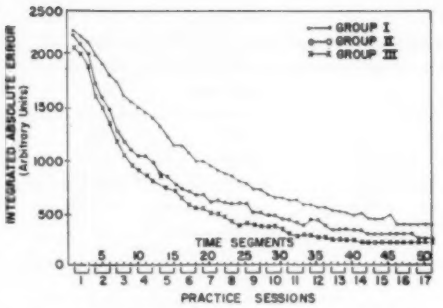

FIG. 2. Comparison of integrated absolute error measures for Groups I, II, and III.

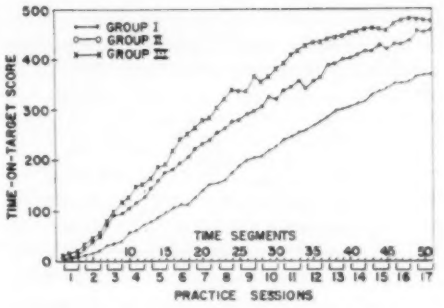

FIG. 3. Comparison of time-on-target measures for Groups I, II, and III.

In terms of each measure there is some indication of the superiority of the experimental procedures over each of the control procedures. Group II (common sense training) appears substantially superior to Group I (no formal training). Group III (experimental training) is superior to Group II but the magnitude of this superiority is not as large as that shown by Group II over Group I.

Analyses of variance (Edwards, 1951) were performed to determine the significance of this apparent effect. Since Group III is clearly superior to Group I, only Groups II and III were compared. This analysis (Table 2) compares over-all group differences during the entire practice period, differences between practice segments, as well as the practice segment-group interaction. A significant interaction would indicate a differential learning *rate* or a different form of learning curve for one group compared with the other.

These analyses indicate that, for both measures of performance, the experimental group is significantly superior to the control group. For the time-on-target scores (Table 3) the Practice Segment × Groups interaction term is significant indicating that one group has either a differential rate of learning or a different form of learning curve. This teraction term is not significant for the integrated error measure, however.

An additional analysis was conducted to determine if the superiority of Group III could be demonstrated using only a measure of terminal performance. Table 4 presents the results. The scores used in this analysis represented the average of the last nine segments of performance for each subject. Thus,

TABLE 2

ANALYSIS OF VARIANCE BASED ON INTEGRATED ERROR SCORES FOR GROUPS II AND III

Source	*df*	*SS*	*MS*	*F*
Group II vs. Group III	1	208,180	208,180	6.25*
Error (*Ss* in groups)	119	3,960,754	33,284	
Time Segments	50	14,471,481	289,430	486.44**
Segment × Group Interaction	50	13,701	274	.46
Error (*Ss* × segments within groups)	5,950	3,537,650	595	
Total	6,170	22,191,766		

* Significant at .05 level.
** Significant at .01 level.

TABLE 3

ANALYSIS OF VARIANCE BASED ON TIME-ON-TARGET SCORES FOR GROUPS II AND III

Source	*df*	*SS*	*MS*	*F*
Group II vs. Group III	1	21,255	21,255	4.64*
Error (*Ss* in groups)	119	544,131	4,573	
Time Segments	50	1,320,240	26,405	546.70**
Segment × Group Interaction	50	5,767	115	2.39**
Error (*Ss* × segments within groups)	5,950	287,160	48	
Total	6,170	2,179,209		

* Significant at .05 level.
** Significant at .01 level.

the score of each individual represents 54 minutes (nine segments times 6 minutes per segment) of tracking performance. This expanded measure was used in an attempt to obtain a stable index of terminal performance for each subject. In terms of integrated error there is a significant indication of the superiority of the experimental procedures. For the time-on-target measures the superiority of the experimental procedures approaches statistical significance. Inasmuch as the variances of Group II and Group III were heterogeneous with respect to both measures, the *t* tests were run in accordance with the Cochran-Cox method as presented by Johnson (1949, p. 75). This procedure takes into account the heterogeneity of variances.

From the above results it appears that the experimental training procedures operated both to improve group performance and to decrease group variability.

TABLE 4

RESULTS OF TESTS COMPARING GROUP II AND GROUP III WITH RESPECT TO TERMINAL PROFICIENCY

	Mean	*SD*	Difference	*t*
Integrated Error (average of last nine segments)				
Group II	306.6	229.1	85.5	2.65*
Group III	221.1	104.1		
Time-on-Target (average of last nine segments)				
Group II	437.3	113.3	33.0	1.77
Group III	470.3	91.0		

* Significant at .05 level.

The analysis presented in Table 4 indicates the experimental training procedures to be superior in terms of resulting terminal proficiency. However, this conclusion is based primarily upon use of over-all performance (integrated error and time-on-target) measures. The integrated error measure in turn is a function of three component scores: azimuth error, elevation error, and sideslip error as related in Equation I. The question then presents itself: Is the conclusion concerning the superiority of the experimental procedures dependent upon a particular component score? That is, are one or more of the component scores producing the major portion of the differences between the groups?

In order to answer this question subjects' scores on each of the three separate component measures were examined. For this purpose, scores during the final nine segments of practice were summed. Table 5 presents the results of *t* tests comparing Groups II and III with respect to each component score. Inasmuch as the variances of Group II and Group III were heterogeneous with respect to each component score, the *t* tests again were run in accordance with the Cochran-Cox method as presented by Johnson (1949).

Table 5 indicates that for two of the component measures (elevation and sideslip error) Group III is significantly superior to Group II. For azimuth error the difference approaches significance ($p = .07$). It appears then that for all component measures the experimental training procedures operate to

TABLE 5

RESULTS OF TESTS COMPARING GROUP II AND GROUP III WITH RESPECT TO COMPONENT PERFORMANCE MEASURES

	Mean	s	Difference	t
Azimuth Error Score (average of last nine segments)				
Group II	691.6	591.7	158.8	1.88
Group III	532.8	292.8		
Elevation Error Score (average of last nine segments)				
Group II	371.2	380.6	178.8	3.43**
Group III	192.4	145.9		
Sideslip Error Score (average of last nine segments)				
Group II	472.8	266.6	187.5	4.90**
Group III	285.3	134.1		

** Significant at .01 level.

improve performance and to decrease variability. There is no indication that the superiority of Group III, as indicated by the integrated absolute error measure, is dependent mainly upon improvement in one particular component measure. Each component task activity contributes to the overall performance superiority of Group III.

DISCUSSION

This study investigated whether certain kinds of analytical information concerning task requirements might be used to increase the efficiency of a training program. The data of this study support the conclusion that training procedures based upon analytical information concerning task requirements can increase the effectiveness of training.

Ambiguous results were obtained with respect to the question of a different rate (and possibly a different form) for the learning curve of the experimental versus that of the control group. With analysis of variance procedures the time-on-target measure indicates a significantly faster rate of learning for the experimental group. The analysis based on the integrated error measure did not confirm these results. In general, however, it is believed that the experimental training procedures used in this study result primarily in an increase in the slope of the learning curve but not in its basic character.

It should be noted that no experimental procedures were introduced after Session 11. Groups II and III were handled in an identical manner during the subsequent days represented by the terminal proficiency measures. In other words, these last trials may be regarded as trials during which transfer effects of the experimental procedures might be evaluated. The results indicate that withdrawal of the supplementary training instructions did not result in a decrease in performance for the experimental group. We may infer, therefore, that changes in *learning* occurred (with consequent transfer to later performances) rather than just changes in *performance* in the presence of these experimental procedures during training.

The fact that the experimental training procedures resulted in a superiority for Group III has certain additional implications. It will be recalled that much of the experimental training procedures were derived from information concerning the importance of underlying ability factors at different stages of mastery of the tracking task. The fact that emphasis of these abilities at appropriate points within the training cycle did result in an improvement in training would seem to bear certain evidence concerning abilities identified through factor analytic procedures. These abilities may be more than just descriptive categories. They may, in fact, represent some mediating process within the organism.

The integrated error measure indicates a 39% improvement in training effectiveness at the terminal stages of training. Yet the original analysis of the tracking task upon which the experimental training procedures were based was not as fruitful as originally hoped. Of the 15 ability factors defined by a factor analysis of the battery of reference measures, only 2 showed a systematic relation to tracking performance. Therefore, only these 2 could be used in structuring the experimental training program. As further research into the nature of the abilities required by tracking tasks or other tasks of similar complexity is accomplished, it should be possible to establish training programs still more effective than used in this study.

SUMMARY

The objective of this study was to investigate the extent to which the effectiveness of training for a complex perceptual-motor activity might be increased through use of training procedures based upon a detailed analysis of task requirements. Specifically, a realistic tracking task was analyzed in terms of the ability factors which were important contributors to proficiency at different stages of practice and the relation of component task performances to terminal over-all proficiency.

Three groups of subjects were used. The three groups were trained on the same tracking task and for identical periods of time. Group I ($N = 203$) received no formal training other than the answering of questions. Group II ($N = 60$) received a "common sense" training program using standard pedagogical techniques. Group III ($N = 60$), the experimental group, received the same training as Group II plus a set of standard instructions derived from the analysis of task requirements.

The following conclusions are drawn:

1. There is an indication of consistent superiority for the experimental procedures throughout the course of practice. These procedures operate both to increase proficiency and to decrease within-group variability.

2. There is no evidence that the experimental program operated differentially upon component task performances such as azimuth error, elevation error, or "sideslip" error.

3. Results indicate that the slopes of the learning functions differed, but this may not mean the basic form of the learning curve was different for each group.

4. The superiority of the experimental group persisted in later tracking sessions, when the experimental instructions were no longer introduced.

REFERENCES

DUBOIS, P. H. The design of correlational studies in training. Paper presented at Symposium on Training Research and Its Implications for Education, University of Pittsburgh, 1960.

DUBOIS, P. H., & MANNING, W. H. (Ed.) Methods of research in technical training. Technical Report No. 3, May 1957, Washington University, ONR 816(02). (A report of a conference held at the Naval Air Station, Memphis; comments by Wilse B. Webb.)

EDWARDS, A. L. *Experimental design in psychological research*. New York: Rinehart, 1951.

FLEISHMAN, E. A. A comparative study of aptitude patterns in unskilled and skilled psychomotor performances. *J. appl. Psychol.*, 1957, **41**, 263–272.

FLEISHMAN, E. A. Description and prediction of perceptual-motor skill learning. Paper presented at Symposium on Training Research and Its Implications for Education, University of Pittsburgh, 1960.

FLEISHMAN, E. A., & FRUCHTER, B. Factor structure and predictability of successive stages of learning Morse code. *J. appl. Psychol.*, 1960, **44**, 97–101.

FLEISHMAN, E. A., & HEMPEL, W. E. Changes in factor structure of a complex psychomotor task as a function of practice. *Psychometrika*, 1954, **19**, 239–252.

FLEISHMAN, E. A., & HEMPEL, W. E. The relation between abilities and improvement with practice in a visual discrimination reaction test. *J. exp. Psychol.*, 1955, **49**, 301–312.

FRUCHTER, B. *Introduction to factor analysis*. New York: Van Nostrand, 1954.

GAGNÉ, R. M., & BOLLES, R. C. A review of factors in learning efficiency. In E. Galanter (Ed.), *Automatic teaching: The state of the art*. New York: Wiley, 1959.

JOHNSON, P. O. *Statistical methods in research*. New York: Prentice-Hall, 1949.

MCGEHEE, W. Are we using what we know about training? Learning theory and training. *Personnel Psychol.*, 1958, **11**, 1–12.

PARKER, J. F., & FLEISHMAN, E. A. Prediction of advanced levels of proficency in a complex tracking task. *USAF WADC tech Rep.*, 1959, No. 59–255.

PARKER, J. F., & FLEISHMAN, E. A. Ability factors and component performance measures as predictors of complex tracking behavior. *Psychol. Monogr.*, 1960, **74**(16, Whole No. 503).

(Received September 2, 1960)

Journal of Applied Psychology
1961, Vol. 45, No. 5, 303–308

OPERATOR PERFORMANCE ON A CHORD KEYBOARD

H. C. RATZ AND D. K. RITCHIE [1]
University of Saskatchewan *Ferranti-Packard Limited, Toronto*

The man-machine link has received considerable attention both in simple feedback systems (Hick, 1952) and as a data processing unit. In a computer system, the human operator may be employed because of his ability to translate between plain language and machine language, while in a control system his ability to recognize complex patterns in performance data is exploited (Elkind & Forgie, 1959).

The performance of an operator in choice and sorting problems has been treated by several investigators (Crossman, 1953). They have found the "choice reaction time" to be proportional to the stimulus complexity as measured by the logarithm of the number of alternatives. Experimental variations have included frequency unbalance among the alternative stimuli (Hyman, 1953), and the addition of extraneous information not relevant to the required choice (Archer, 1954; Gregg, 1954). Fitts (1953) has pointed out that an operator's data processing capability is subject to the fixed constraints of his motor system. These constraints are implicit in the form of the response code and determine his effective capacity. For example, a telegraph operator produces time sequences with one hand, while the teletypist uses 10 fingers one at a time. The fact that motor constraints have a considerable influence in the latter case is known from the improvements that can be achieved through keyboard and language studies (Dvorak, Merric, Dealey, & Ford, 1936). The type of keyboard studied in this paper is more akin to that used by the stenotypist who employs several fingers simultaneously. At present, the typewriter remains the most effective man-computer communication link (Licklider, 1960) although other aspects of computer systems such as displays and programing language have advanced considerably. The multiple finger or "chord" keyboard may be a useful alternative in some special purpose data processing systems.

The relative difficulty of the various chords can be measured in terms of the reaction time in responding on the keyboard to a visual presentation of the chord pattern. Since lights, fingers, and keys are in direct correspondence, the stimulus-response codes are highly compatible; that is, mental recoding of the information is avoided. Thus we obtain an experimental assessment of the performance of an operator using this chosen set of response motor tasks. Using these data and the principles of optimum coding, the more frequently used message units can be assigned to the easier response tasks. This optimum distribution of message units will minimize the average time per message and maximize the rate of information processing (Blachman, 1954).

METHOD

Apparatus

A block diagram of the experimental apparatus is shown in Figure 1. A paper tape reader provided sequences of patterns for the lamp display. The operator responded to each pattern in turn by depressing the corresponding keys of the keyboard. A complete temporal record of both the display and response activity was obtained from a 20-pen recorder. Ten of the pens recorded the motions of the keyboard responses, while the other 10 recorded the timing of the lights. Thus reaction time, errors, latency, etc., could all be obtained from the paper strip record.

The keyboard contained 10 keys mounted on a horizontal surface and placed at the positions occupied by the fingertips with the fingers slightly curled and the wrists straight. The keys operated snap action switches through a lever which required about three-eighths inch travel at the key, and quite low pressure. These switches drove the recording pens and, when all keys were released, signaled the tape reader to present the next stimulus pattern. The lights used in the stimulus display were placed in two groups of five on a horizontal line above and beyond the keyboard. They were mounted on a surface tilted so as to be roughly perpendicular to the line of sight.

[1] The authors gratefully acknowledge the assistance and guidance of John Ogilvie and M. Humphries in the formulation of these experiments and in the analysis of the data. The valuable comments on the manuscript by George A. McMurray are also appreciated.

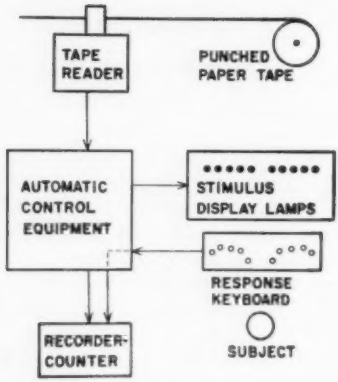

FIG. 1. Block diagram of the experimental apparatus. (The flow of stimulus and response information is indicated. Control signals are omitted.)

In a typical experiment, the operator depressed the keys on the keyboard which corresponded to the combination of illuminated lights presented to him. That is, the operator transcribed light "chords" into key "chords." When all keys were released by the operator, a new combination of lights would appear whether or not the correct chord had been struck.

The function of the control apparatus was to automatically step on the tape reader and to transfer the combinations punched on the tape to the lamps. The method of operation as described above is one in which the operator paces himself; that is, the new stimulus was automatically presented after the keys were released from the previous response. Although not used in these experiments, forced pacing at various speeds, and with a choice of ratio of "on time" (stimulus presented) to "off time" (dead time between stimuli), could be programed with the same control apparatus.

Stimuli

The tape reader was a standard type used with computers and employed five-hole tape. For experiments using chords in only one hand, successive characters on the tape provided the stimulus combinations. By using a pair of characters from the tape for each presentation, a full 10-light pattern was achieved. This choice of program could be preset at the control apparatus. Thus when both hands were used the reader passed two characters of the tape at a time.

Since the sequence of light patterns depended only on the characters of the punched tape, many experiments with various sets of chords could be easily programed. For example, one tape employed all the 31 different possible combinations with equal frequency but in random order; another employed only those 15 different chords per hand which are possible using either one or two fingers.

Subjects

Six operators were used in the experiments. Three of these used both hands throughout, while the other three used only the hand of preference, which was the right hand. Each operator practiced 10 minutes a day on random sequences of the 31 chords. The performance was completely recorded for the first and last 100 seconds of each 10-minute interval.

Although the task required in these experiments is a simple motor reaction, there is an initial short learning period. By checking the average reaction time of a few of the chords, it was observed that little improvement took place after the second day. Asymptotic behavior was reached rather quickly because the order of the presentations was random. The absence of any systematic intersymbol influence between successive chords precluded any longer term improvement through learned predictability since, in this case, each of the possible stimuli was equally likely to occur at any point in the sequence. In other terms, the stimuli signals contained no redundancy, so the gross data rate and the net rate of informational entropy were equivalent. Finally, because self-pacing was used, the number of erroneous responses made was negligible.

Procedure

In order to rank the chords according to increasing reaction times, a stimulus tape using all 31 different chords at random was used to generate patterns. The daily test periods and recording intervals were the same as during training. From the total of 2,660 measured reaction times, each chord was assigned a rank number according to its position in a list of increasing reaction times. The reaction time was measured between the appearance of the lamp stimulus pattern and the complete selection of the corresponding chord. The data for each chord was totaled for the complete duration of the experiment since an operator's behavior might vary 10% from day to day.

In the second part of the experiment, information rates were measured using selected groups of stimulus chords. Four groups were defined which used a maximum of one, two, three, or all five fingers per hand. They appeared in random sequence with equal frequency. The appropriate information measure is the logarithm of the number of possible choices, which is the average entropy per pattern of the source, i.e.:

$$H = \log_2 (\text{number of choices})$$

The information rate (bits per second) is obtained from the product of H and the average number of responses per second. The latter was measured by counting the total number of responses in 3-minute tests for the cases involving only one hand, and in

RANK	CHORD	RANK	CHORD
1	– – – – O	17	O – O – –
2	O – – – –	18	O – – O –
3	– – O – –	19	– – O O O
4	– – – O –	20	O – – O O
5	– O – – –	21	O O – – O
6	O O – – –	22	O – O – O
7	O – – – O	23	– O – – O
8	– O O – –	24	O – O O O
9	O O O – –	25	– – O – O
10	– O O O –	26	O – O O –
11	– – O O –	27	O O – O –
12	– – – O O	28	O O O – O
13	O O O O O	29	O O – O O
14	– O O O O	30	– O O – O
15	O O O O –	31	– O – O O
16	– O – O –		

FIG. 2 Chord rank chart. (The fingers used in the right hand are indicated.)

6-minute tests for the cases with two hands. The six different experiments are listed in Table 1.

RESULTS AND DISCUSSION

Chord Ranking

The average ranking of chords by increasing reaction times is shown in Figure 2, where the fingers used in each chord are indicated for the right hand. The average correlation coefficient among operators for all 31 chords is .91, hence, the agreement on the overall ranking is excellent. However, there is no significant agreement among operators concerning the rank of any chord relative to those immediately adjacent on the list. For example, if we confine our attention to the first five on the list, the correlation in the ranking by different operators is not significant; but if eight or more chords are taken the agreement is good. This is clear from the maxima and minima for the different subjects in Figure 3.

The main features of the chord rank chart are not unexpected, thus, the top five on the list involve only one finger each. The chords requiring the longest response times all contain patterns requiring one or two fingers to be held off while their neighbors on each side are used. In between (ranks 8 to 15) lie those chords consisting of a consecutive group of fingers.

Distribution of Response Times

If each chord is assigned a "cost" proportional to its associated reaction time, then the distribution of Figure 3 is obtained. These response times include a 0.1-second delay in the presentation of the stimulus pattern. In passing, it was observed that about 65% of the total time required to complete the chord response could be assigned to latency, where latency is taken as the time between stimulus presentation and the first indications of response. Figures 2 and 3 give the relative ranking of chords by response time and show quantitatively how the cost in reaction time is distributed over the different chords.

The quantitative difference in performance using various subgroups of chords can be predicted from Figure 3. If it is assumed that the cost distribution of Figure 3 represents the motor response times of each chord and that the effect of choice when there are different numbers of possible alternatives is negligible, then the expected response time would be the average cost (in seconds of reaction time) determined from the distribution for the given subgroup of stimulus patterns. The "predicted response time" of Table 1 is the average of the costs of those chords used in a particular group of stimulus patterns plus the 0.1-second dead time between presentations. To extend

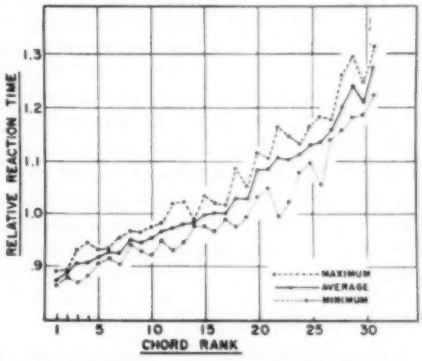

FIG. 3. Distribution of reaction times. (Chords are ranked according to Figure 2 and reaction times are expressed relative to the median of 1.16 seconds.)

TABLE 1

COMPARISON OF THE OBSERVED DATA RATE FOR SUBGROUPS OF CHORDS WITH THAT PREDICTED BY THE CHORD RANKING EXPERIMENT

Experiment	Stimulus Chords	Patterns H (bits/stimulus)	Observed Response Time T seconds/ response)	Observed Data Rate H/T (bits/sec.)	Predicted Response Time C (seconds/ response)	Predicted Data Rate H/C (bits/sec.)
One Hand						
A	1-finger chords	2.32	.94	2.4	1.05	2.2
B	1-, 2-finger chords	3.91	1.07	3.7	1.13	3.5
C	1-, 2-, 3-finger chords	4.64	1.15	4.1	1.18	3.9
D	All chords	4.95	1.20	4.1	1.20	4.1
Two Hands						
E	1 finger per hand	4.64	2.08	2.3	(2.1)	(2.2)
F	All chords	9.91	2.63	3.8	(4.1)	(4.1)

this to two hands, the distribution for all 961 allowable 10-finger patterns would be required. Lacking these data, the figures entered in parenthesis in the table are merely double the expected reaction times for one hand.

Table 1 shows the results of experiments using only selected groups of chords and measuring the average response time for the group in each case. The "Observed Data Rate" is obtained by dividing the H for the group by the observed average response time T. Since H measures the stimulus information in binary units, the result is an information flow rate expressed in bits per second which neglects the possibility of errors. As is usual in this type of human operator experiment, the higher information rates correspond to the stimuli giving the greater choice and having higher values of H. While this conclusion is valid here for the cases involving one handed operation, note that it does not necessarily extend to the use of two hands. Comparing Experiment A with E, and D with F, it is seen that an operator using only one hand may perform at somewhat better than twice the speed of one using both, and therefore having a greater resulting information handling capacity. Thus the increase in choice derived from using both hands on 10 keys was more than offset by the slower response of the operator.

Information Rates

The agreement between the observed response time T and the expected average cost C indicates that the reaction time for a given subgroup of chords is very nearly the average of the reaction times for its individual members. Thus the assumption made in computing C is substantially supported, namely, that the reaction times are highly characteristic of the individual chords and relatively independent of the number and remaining members of the set. However, there is a tendency for the response times to be shorter (and the information rate higher) than that predicted when the amount of choice in the set becomes smaller; that is, the effect of the amount of choice on the overall response time may be present but is secondary to the average motor reaction time. Since the basis of all predictions was reaction time measurements using all 31 chords, the agreement in Experiment D is an arithmetic check.

In Table 1, the highest data rate is obtained using all chords on one hand. However, this assumes that the 31 different stimuli patterns are used with equal frequency, a method which is not optimum in view of the nonuniform cost distribution function. Obviously, the more difficult chords should be used less often than the easier ones. The frequency distribution required for optimum performance can be determined from coding theory.

Application of Coding Theory

The problem of communication with unequally weighted vocabularies has been examined as a game of strategy by Mandelbrot (1954). By interpreting the average response time to each of the chords as its cost, the code can be specified which will transmit a given amount of information in minimum time. Under the assumption that the response time to

a chord is independent of its predecessors or the amount of choice, the encoding which minimizes the cost can be essentially specified by the relative frequencies of use of the chords (Blachman, 1954). Since Table 1 shows this assumption to be approximately true, the maximum information rate will be achieved by assigning the easier chords to the more frequent messages.

Suppose the chords are given a rank $R = 1, 2, \ldots, 31$ as shown in Figure 2 with an associated cost (C_R) from the experimental results of Figure 3. The problem is to find the relative frequency of use for each chord—its probability p_R—that will maximize the net information rate H/C. Here H is the weighted average entropy per chord:

$$H = -\Sigma p_R \log_2(p_R)$$

and C the average cost:

$$C = \Sigma p_R C_R$$

The probabilities are normalized so that:

$$\Sigma p_R = 1$$

The solution of this variational problem is in terms of the "partition functions" of statistical mechanics (Jaynes, 1959); the distribution is of the form:

$$p_R = 2^{-kC_R}$$

where k, the maximum information rate, is the solution of the equation:

$$\Sigma 2^{-kC_R} = 1$$

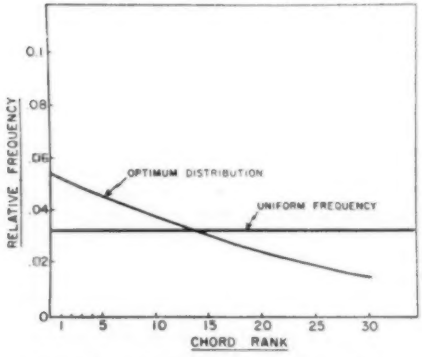

FIG. 4 Comparison of uniform with optimum frequency distribution.

Using Figure 3 as the description of C_R, the information rate obtained with optimum frequency distribution of the chords is:

$$k = 4.33 \text{ bits/second}$$

The resulting distribution is compared in Figure 4 with the uniform frequency method of Experiment D. Note that the frequency of occurrence of the most used chords ($R = 1$) is about 3.7 times that of the least used ($R = 31$), with the unbalance in favor of the easier chords. The improvement to be expected over the results of Experiment D is of the order of 5% in the net information rate, if the frequency of use is optimum as shown in Figure 4.

SUMMARY

The 31 chords have been ranked according to their difficulty or "cost" as measured by the reaction times, and a quantitative measure obtained for the distribution. Assuming that the motor system predominates over choice reaction time, these results are used to predict the expected average response times and information rates using subgroups of the chords that involve less choice; for example, a group may include only a few of the easier chords with the shorter response times. The agreement with experiment indicates that the results were determined primarily by the reaction times of the particular chords in the group, and that the effect of the amount of choice was secondary. On this basis, coding theory can be used to deduce the distribution of the frequency of occurrence of chords with the measured cost function that gives the maximum information rate through the man-machine link.

For one hand the information rate increases with the complexity of choice of patterns, but this does not remain valid for two hands where the loss in response speed overbalances the increased choice of stimuli. It should be noted that these keyboard experiments were conducted using a highly compatible display, and that other forms of presentation might introduce effects of stimulus-response coding that would reduce the importance of the motor reaction time. With this qualification, the results show that motor system constraints are predominant over choice reaction time in determining speed on a chord keyboard.

REFERENCES

ARCHER, E. J. Identification of visual patterns as a function of information load. *J. exp. Psychol.*, 1954, **48**, 313–317.

BLACHMAN, N. M. Minimum-cost encoding of information. *IRE Trans.*, 1954, **IT-3**, 139–149.

CROSSMAN, E. R. F. W. Entropy and choice time: The effect of frequency unbalance on choice response. *Quart. J. exp. Psychol.*, 1953, **5**, 41–51.

DVORAK, A., MERRIC, N. I., DEALEY, W. L., & FORD, G. C. *Typewriting behavior*. New York: American Book, 1936.

ELKIND, J. I., & FORGIE, C. D. Characteristics of the human operator in simple manual control systems. *IRE Trans.*, 1959, **AC-4**, 44–55.

FITTS, P. M. The influence of response coding on performance in motor tasks. In, *Current trends in information theory*. Pittsburgh: Univer. Pittsburgh Press, 1953. Pp. 47–75.

GREGG, L. W. The effect of stimulus complexity on discrimination responses. *J. exp. Psychol.*, 1954, **48**, 289–297.

HICK, W. E. On the rate of gain of information. *Quart. J. exp. Psychol.*, 1952, **4**, 11–26.

HYMAN, R. Stimulus information as a determinant of reaction time. *J. exp. Psychol.*, 1953, **45**, 188–196.

JAYNES, E. T. A note on unique decipherability. *IRE Trans.*, 1959, **IT-5**, 98–102.

LICKLIDER, J. C. R. Man-computer symbiosis. *IRE Trans.*, 1960, **HFE-1**, 4–11.

MANDELBROT, B. Simple games of strategy occurring in communication through natural languages. *IRE Trans.*, 1954, **IT-3**, 124–137.

(Received September 14, 1960)

Journal of Applied Psychology
1961, Vol. 45, No. 5, 309–317

NOISE, THE "AROUSAL HYPOTHESIS," AND MONOTONOUS WORK[1]

WILLIAM N. McBAIN

San Jose State College

What is the nature of the relation between noise and performance? Over the past 45 years there has been a discouraging lack of agreement among the proposed answers to this question. Berrien (1946) concluded that findings were at best "contradictory," while Melton and Briggs (1960) characterized results from the previous 2 years' research as "suggestive, but not easily integrated." Indeed, we have advanced little toward a general answer since Morgan (1916) first raised the question.

The present study was undertaken in an attempt to clarify the problem by placing it in the larger framework of a theory dealing with the effects of environmental stimulation, generally. It combines the recent trend toward specifying the kind of work and of noise involved with a type of dimensional analysis of the particular classes of work and noise used. Though the experimentation is guided by these theoretical considerations, its focus is upon the practical problem of reducing errors in industrial output and of selecting personnel least liable to such errors.

Noise and Arousal

An attribute possessed by noise, in common with other stimuli, is *variability*. Very rarely, except in a laboratory situation, is an organism exposed to a constant, unvarying sound—*change* in stimulation is the rule. It has been proposed (Hebb, 1955) that changes in stimulation not only serve as cues to the organism, but also to activate or *arouse* areas of the cortex which are involved in the effective use of (or reaction to) such environmental cues. Marked impairments in coping behavior have been found (Bexton, Heron & Scott, 1954; Scott, Bexton, Heron, & Doane, 1959) when environmental stimulation has been severely restricted.

The theoretical relation between effectiveness and the degree of arousal is not a linear one, but may be represented as an inverted U when the former variable is plotted on the vertical axis and the latter on the horizontal. For a specific person, performing a particular job, a given degree of arousal should lead to optimal performance, while lower or higher arousal levels will be associated with reduced effectiveness. Since all sensory inputs are routed to the nonspecific arousal system, any environmental change, including changes in auditory stimulation, should result in increased arousal.

While the degree of arousal is chiefly a function of, and dependent upon, the total variablity of the stimulus situation, there seems also to be a feedback situation which allows higher mental processes such as problem solving or imagining to contribute. Though these internal sources of arousal are apparently insufficient to maintain optimum effectiveness, our knowledge of them requires that they be controlled in experimentation.

Noise and "Intelligibility"

There is a further characteristic of noise which, though it has little to do with either variability or intensity, seems to have a great deal to do with the extent to which a noise will prove distracting. "Distracting" is used here in the sense of successful competition for attention and is demonstrated when loud shouts in the hall spoil the rapport between

[1] The research reported in this paper was supported by the Defence Research Board of Canada, under Grant No. 9401-12, Project D77-94-01-12. The author, a faculty member of McGill University at the time, wishes to acknowledge the assistance of Alan Lyall, who not only aided in the planning and analysis, but also performed most of the experimental work on a very strenuous schedule. Special thanks are due to F. S. Carpenter, Air Officer Commanding, Transport Command, Royal Canadian Air Force, who made available the experimental facilities, and to members of his staff.

students and lecturer. For want of a better name this characteristic is referred to as "intelligibility," which implies that it is a joint function of the stimulus *and* the person stimulated. Roughly, a noise may be considered intelligible to the extent that it carries meaning within the frame of reference of the individual, rather than being perceived simply as a noise. While languages may exhibit this characteristic most clearly, the squeal of a dry bearing or the ring of a coin falling on pavement possess a high degree of intelligibility for appropriate auditors.

In experimental situations one can control the intelligibility of the noise used by holding it at approximately zero (pure tones, white noise, etc.), or by keeping it at a low level which can be presumed to be approximately equal for all subjects. Music used in work situations is ordinarily restricted to instrumental, and would be considered an example of the second approach. To the extent that sound is intelligible it can be presumed to be distracting, and hence likely to produce errors in the performance of a task demanding continuous attention.

Monotonous Work

When applied to a work situation the two terms "boring" and "monotonous" have frequently been used as synonyms. In this study the term boring will have the usual reference to certain unpleasant subjective states often associated with the performance of jobs in which one has no particular interest. Monotonous, on the other hand, will be reserved to refer to a *total* work situation which offers little variability, and hence little opportunity for arousal. Monotonous, when used in this way, is also a characteristic of the stimulus situation and the indvidual. The work situation is monotonous to the extent that it lacks variability, which is a relatively objective and (at least conceptually) quantifiable characteristic. Since it is not possible to quantify the relatively less important sources of variability that come from the individual's imaginative and cognitive processes, an experimental task should be such as to hold these at the lowest possible level.

To be specific, in terms of the foregoing considerations a task might be said to be monotonous to the extent that it (*a*) involves in itself and the total work situation as little variability or stimulus change as possible—variability should be repetitious to the extent that it does occur; (*b*) demands the individual's attention as continuously as possible, so that imagination and fantasy systems are unlikely to develop; (*c*) is sufficiently structured and planned that a minimum of cognitive activity is required in its performance.

The laboratory task developed to meet these criteria required each subject to hand print continuously, at a paced rate, a sequence of seven pairs of letters. In comparison to most procedures used to test "perceptual vigilance" (Broadbent, 1953), this task has the advantage of having the subject continuously and actively involved, so that lapses of alertness can be traced to the exact time of occurrence. It also resembles more closely the machine paced conditions of performance which characterize many production line jobs.

Arousal and Hypnotic Susceptibility

Long distance truck driving, especially at night, is a job which comes close to meeting the criteria of monotony given above. The Harvard School of Public Health study of *Human Factors in Highway Transport Safety* (McFarland & Moseley, 1954, p. 124–125) points out that hallucinations are a recognized hazard of this job. The isolation studies (Bexton et al., 1954) showed that reducing stimulus input frequently leads to various types of hallucinations. Ammunition inspectors have been reported as "falling asleep" at this highly repetitious visual task, frequently relatively early in a shift.[2] The resemblance of these occurrences to phenomena which are common in the hypnotic state suggested a relation between the individual's hypnotic susceptibility and his reaction to a low variability stimulus situation. Though this suspected relation may be tangential to the effects of noise, its investigation was thought likely to add to the understanding of behavior in a monotonous work situation.

[2] E. C. Webster, personnel communication, 1956.

Experimental Hypotheses

The major experimental hypothesis is implicit in the arousal hypothesis, which predicts lowered efficiency following the reduction of environmental stimulation to low levels. In a monotonous work situation (defined in terms of a very low level of stimulus change) the introduction of noise, by increasing variability and hence arousal, should lead to improved performance.[3] As previously noted, the noise introduced should be low in intelligibility, otherwise its distracting effect may mask any performance improvement attributable to increased arousal.

A second hypothesis asserted that a relation would be found between a standard measure of susceptibility to hypnosis and the amount of performance decrement in the task occurring from conditions of noise to conditions of minimal variability.

An additional purpose of the investigation was to assess the laboratory task itself as a reliable index of interindividual differences in performance under monotonous work conditions. If demonstrated to be reliable, the resulting increase in behavioral sampling should lead to improved validity of selection for such jobs (McBain, 1957).

Method

Subjects. Thirty male members of the Royal Canadian Air Force formed the original group. Their ages ranged from 19 to 36 years with a mean of 25.8. All, except two of the three officers in the sample, were nonflying personnel. Attrition from unscorable records, mechanical failure, and equating of groups reduced the N to 26 or less for most data presented.

Procedure.[4] Each subject completed four 42-minute sessions of work at a monotonous task. For almost all subjects these were spaced at exactly one-week intervals. Prior to these sessions, each subject was administered a slightly modified form of the Friedlander and Sarbin (1938) Scale of Hypnotic Depth, essentially an instrument which gives a score reflecting the individual's reaction or susceptibility to a measured "dose" of hypnotic stimuli.

The Monotonous Work Task. A laboratory task was devised to fit the criteria of a monotonous job, as defined above. This consisted of hand printing continuously in sequence the seven pairs of letters: KT, LH, IM, XF, KV, TZ, IW. Each of the pairs consisted of letters which were easily discriminable for scoring purposes and in ordinary printing each required the same number of strokes. Prior testing demonstrated the sequence to be highly confusing. This, combined with pacing, helped to meet the task criterion of being attention demanding. To reduce even further the cognitive aspects of this highly structured task, the list of letter pairs was posted so that the subject could consult it by briefly looking away from the printing surface.

The requirement of minimum variability in the total work situation was met by several means. Intruding exterior noises were masked to a considerable extent by the continuous operation of a powerful exhaust fan during sessions, so that the only variable sounds in the experimental room were the rhythmic clicks of the timing and recording system. Visual variability was reduced by having the performance carried out in a black draped cubicle which was visually isolated from the rest of the experimental room. Kinesthetic variability was controlled to some extent by cautioning the subject to move about as little as possible.

At the beginning of the first work session each subject was given a pretest which required rapid copying of the letter pairs on a prepared test blank. Three 36-second trials were given. The subject was encouraged to print as quickly as possible. For the first three of the succeeding 42-minute sessions each subject was required to respond at three-fifths of his average rate on the two fastest pretest trials. On the fourth session each subject was required to work at a slower rate.

The pacing apparatus presented to the subject a $7/8'' \times 1''$ opening beneath which ran a continuous tape of white paper. This advanced intermittently at the rate set for each subject, and in such a manner that the previously printed pair was immediately concealed. Each time the subject made an error (defined as "anything in the space except the next pair of letters in the sequence, printed legibly and completely") he was required to signal his awareness of this by means of a foot pedal. If the subject lost his place in the sequence he was required to begin again at the first pair, and was informed that this counted as an error. Subjects were instructed to make as few errors as possible.

Conditions. Three basic conditions were used in the four sessions. Under Quiet (Q) conditions the only sounds in the experimental room were the rhythmic clicks of the apparatus and the masking sound of the fan. During the entire period of the Noise (N) sessions a magnetic tape was played. Its content was considered to be low in intelligibility, since it was basically a recording of speech played in reverse, specifically, part of a labor-management conference. To add to its variability

[3] Scott's (1954) unpublished suggestion to this effect is one of the bases for the experiment reported. Deese (1955) published a similar suggestion at about the same time.

[4] The complete experimental protocol is contained in the project terminal report submitted to the Defence Research Board of Canada.

there were superimposed short fragments of intelligible conversation, music, meaningless sounds, etc., and the loudness was varied unsystematically from short periods of complete silence and near inaudibility to a degree which approached the discomfort level. Each subject was exposed to the same portion of the tape recording at the same intensity. After being told that recorded noise would be used, the subject was instructed:

> Try not to let it affect your work at all, but just work straight ahead as you did before, trying to make as few errors as possible. Some people think this sort of work background relieves the monotony . . . others think it is distracting. However, we'd like you to work straight on, paying as little attention to it as possible.

The final session for all subjects was termed the Latency (L) session. It was the same as the Q session except that (*a*) the speed of response was slowed from three-fifths to one-half the pretest rate and (*b*) an additional response was required of all subjects, which involved a slight modification of the apparatus.

A small lamp was added to the apparatus, positioned about 6 inches beyond the printing aperture from the subject. Next to it was a push button. On 11 occasions during the session, at unsystematic intervals of 2–6 minutes, but the same for each subject, the lamp was turned on at exactly the time of the paper shift. The subject was instructed to "put it out by pushing the button as fast as possible, without either putting down your pen or missing a turn, and before you print the two letters in the square." To make this task more discriminating the 6-volt lamp was supplied from a 3-volt current source, which had the effect of causing it to glow rather than to shine brightly. Each subject was given five practice light signals prior to beginning the scored trials.

All subjects worked under Q conditions during their first monotonous work session. They were then divided into two groups, equated on the basis of proportion of errors and total number of responses required during the first 42-minute session. In the second session one of these groups performed under Q conditions, the other under N conditions. In the third session the conditions of performance were reversed, while in the fourth session all subjects worked under L conditions.

During each session of each subject readings were taken at regular intervals of the level of galvanic skin response. Since these data have little applied interest they are not included in the present report.

Treatment of Data. For each subject the following scores were computed:

1. The Error Index (EI) consisted of the number of response errors for each thousand responses required, and was computed for each session and for the four sessions combined.

2. The Signal Index (SI) was the percentage of errors scored which was signaled by the subject. It was computed for each session, and for the four sessions combined.

3. The Latency Index (LI) was secured from the kymograph records of the final session. The linear distances between "on" and "off" signals for the final 10 of the 11 trials were summated for each subject. Since relative rather than absolute times were required, this index was not converted to a time score.

4. The Hypnotizability Index (HI) was the score on the Friedlander and Sarbin scale. High scores indicate greater, and low scores lesser, susceptibility to hypnotic stimuli. The subjects were considered "high" or "low" on this index as they fell above or below the median for the group.

Results

Noise, Hypnotizability, and Performance

The original experimental plan called for groups matched on Session I to perform on Sessions II and III in a counterbalanced design. With this intention thwarted by perverse apparatus and subjects, it was decided to analyse the results of equated rather than matched groups, using a complex analysis of variance design with three dimensions:[5] Condition (N vs. Q—correlated), Order of Conditions (NQ vs. QN—correlated), Hypnotizability (High vs. Low—uncorrelated).

Since measures under the two conditions were available only for EI and SI scores, separate groups were equated on these two measures for the first session, which all subjects performed under the same (Q) conditions. Tables 1 and 2 demonstrate that these groups were not significantly different. The results of Sessions II and III were then

[5] Cynthia Wimer planned and performed the analysis presented.

TABLE 1

Equating of Experimental Groups on Error Index (EI) Performance in Session I

($N = 24^a$)

Source of Variation	*SS*	*df*	*MS*	*F*
Hypnotizability	42	1	42	<1
Order (of N & Q conditions)	95	1	95	<1
Interaction (H × O)	799	1	799	2.64
Error	6,062	20	303	
Total	6,998	23		

[a] Subjects 5 and 22 omitted.

TABLE 2

EQUATING OF EXPERIMENTAL GROUPS ON SIGNAL INDEX (SI) PERFORMANCE IN SESSION I

($N = 26$)

Source of Variation	*SS*	*df*	*MS*	*F*
Hypnotizability	720	1	720	1.06
Order (of N & Q conditions)	40	1	40	<1
Interaction (H × O)	2,128	1	2,128	3.14
Error	14,908	22	678	
Total	17,796	25		

analyzed, with the results shown in Table 3 (groups equated for EI) and Table 4 (groups equated for SI).

In brief, the results from these analyses indicated:

1. The difference in mean EI between N conditions (12.22) and Q conditions (16.39) of 4.17 errors per thousand required responses is significant beyond the 5% level of confidence. As measured by the *making* of errors, performance under conditions of noise is superior to performance under quiet conditions.

2. The difference in mean SI between N conditions (55.13) and Q conditions (56.45) of 1.32% of errors signaled is not significant. As measured by the *signaling* of errors, performance under conditions of noise is not significantly different from that under quiet conditions, though the percentage of errors signaled under quiet conditions is slightly greater.

3. There is a highly significant interaction between Order and Condition in their relation to SI. This can best be understood as a practice effect, i.e., there is a tendency for the signaling of errors to be more accurate under the performance condition experienced second. There is no such effect in regard to the making of errors.

4. There is no difference in mean EI between subjects with low and with high HI. However, the mean SI for subjects with low HI (69.34) and for subjects with high HI (42.24) showed a difference of 27.1 in percentage of errors signaled, which is significant at beyond the 5% level. While persons who are least susceptible to hypnosis make no fewer errors than those who are most susceptible, they do signal the errors made with significantly greater accuracy.

The mean LI was determined in Session IV for subjects low in hypnotizability (29.8) and those high in this characteristic (30.9). The difference of 1.1, in favor of those least susceptible to hypnosis responding more quickly, was not significant.

TABLE 3

ANALYSIS OF VARIANCE OF ERROR INDEX (EI) PERFORMANCE IN SESSIONS II AND III

($N = 24$*)

Source of Variation	*SS*	*df*	*MS*	*F*
Hypnotizability	1	1	1	<1
Order (of N & Q conditions)	333	1	333	<1
Condition (N or Q)	209	1	209	4.66*
H × O	1,781	1	1,781	3.30
O × C	21	1	21	<1
H × C	26	1	26	<1
O × H × C	109	1	109	2.43
Subjects	10,804	20	540.2	12.04**
Residual	897	20	44.8	
Total	14,181	47		

* Subjects 5 and 22 omitted.
* Significant at .05 level.
** Significant at .01 level.

TABLE 4

ANALYSIS OF VARIANCE OF SIGNAL INDEX (SI) PERFORMANCE IN SESSIONS II AND III

($N = 26$)

Source of Variation	*SS*	*df*	*MS*	*F*
Hypnotizability	9,545	1	9,545	6.64*
Order (of N & Q conditions)	245	1	245	<1
Condition (N or Q)	23	1	23	<1
H × O	1,601	1	1,601	1.11
O × C	4,340	1	4,340	9.86**
H × C	1,129	1	1,129	2.56
O × H × C	33	1	33	<1
Subjects	31,606	22	1,437	3.26**
Residual	9,686	22	440	
Total	58,208	51		

* Significant at .05 level.
** Significant at .01 level.

Characteristics of the Performance Indices

It seemed possible that a task of this general type might prove useful in the selection of personnel for jobs of an unavoidably monotonous nature. For this purpose it should demonstrate reasonably wide and con-

sistent differences among individuals. Table 5 indicates that even after more than 2 hours practice on this relatively simple task subjects are reasonably well distributed on both EI and SI. An examination of the table indicates a relatively consistent lowering of the mean errors and narrowing of the spread of scores in the EI which is not shown in the SI. It is thought that the latter observation reflects the inadequacy of the SI as a statistic. As the number of errors becomes small a very small absolute change in the number signaled changes the SI tremendously. For example, a subject having only one error will secure an SI of 100 if he signals it, while his SI is 0 if he fails to signal it.

The reliability of the EI was estimated for Session I by the split-half method. The errors for successive 2-minute periods (with the exception of the first 2 minutes) were summated so that errors for Minutes 3, 4; 7, 8; 11, 12; etc. were correlated with errors from Minutes 5, 6; 9, 10; 13, 14; etc. This gave a coefficient of reliability of .84, which increased to .92 when the Spearman-Brown correction for length was applied.

Since it is possible that errors in one 2-minute period may influence errors in the successive one, and hence lead to a spuriously high coefficient, the errors in session II were correlated with those in session III. While it is true that half the subjects worked under different conditions in Sessions II and III, it may also be assumed that, even under N conditions the task is a monotonous one, though not as monotonous as under Q conditions. For 24 subjects errors in Session II vs. errors in Session III showed a correlation of .84, a reasonable finding for a test-retest reliability after one week, considering both the disparate conditions and the restriction of range following Session I. Both these findings indicate a sufficiently high degree of reliability for the EI as a performance measure to make its use feasible as a predictor in an industrial situation.

TABLE 5

SCORES FOR 26 *Ss* IN FOUR MONOTONOUS WORK SESSIONS

Session	Error Index (EI)				Signal Index (SI)			
	M	σ	Extreme Scores		*M*	σ	Extreme Scores	
			Lowest	Highest			Lowest	Highest
I	30.2	21.9	0	97.4	50.4	25.2	0	100
II	15.9	16.8	0	58.2	46.6	28.6	0	100
III	14.2	19.1	0	91.0	61.8	30.1	0	100
IV	12.0	10.6	1.2	42.1	52.5	31.1	0	100

The reliability of the SI, determined as a test-retest coefficient between sessions II and III, was found to be .56, a result most probably attributable to the instability of the index when small numbers of errors are made. From this result it is clear that the SI computed in this manner is not an adequate performance measure for prediction in a personnel situation, even though it has been shown to discriminate between groups selected on the basis of hypnotizability.

The reliability of the LI was estimated by summating the latencies for signals given at Minutes 3, 9, 17, 33, and 41, and at Minutes 5, 15, 23, 27, and 35 for each subject and correlating the resulting pairs of totals. Each of the two sums contained two latencies secured from signals given after 2 minutes of uninterrupted work, one from a signal given after 4 minutes, and two from signals given after 6 minutes, though these intervals occurred in different orders in the two groups of signals summated. The resulting split-half reliability, corrected for length, was found to be $r = .85$, which is of a magnitude usually considered sufficient for use in prediction as well as for comparison of group results.

The interrelation among indices was determined by computing correlations which used mean measures for all four sessions for the EI and the SI and the single determination of LI from Session IV. The three measures were all intended to be indications of the "alertness" of subjects. Thus, although they could be considered relatively independent in terms of experimental procedures, it was anticipated that they would show a reasonable degree of interrelation. For 26

subjects EI was found to correlate with SI .28 and with LI .04 while LI and SI also showed a correlation of .04. Since a correlation of .388 is required for significance at the 5% level with 26 pairs of scores, the hypothesis implicit in the selection of these measures of alertness is clearly untenable. The findings suggest that alertness is not a unitary characteristic in work behavior, and that persons who can be depended upon to perform adequately on a routine, repetitive task which requires continuous attention (EI) are not necessarily those who will be most critical of the adequacy of their performance (SI) or who will react most quickly to an irrelevant interjected stimulus (LI).

Discussion

Arousal and Vigilance Tasks

The majority of recent studies concerned with the effects of noise on human performance have secured data from tasks of "vigilance," in which the subject typically had to respond "only to very infrequent signals but may have to watch for them over long periods" (Broadbent, 1958, p. 108). Two other features are characteristic of the studies summarized by Broadbent. They are primarily concerned with decrements in performance over the length of the period during which exposure to noise occurs. More important to the present study, they usually use relatively homogeneous noise, specified as to frequency band or spectrum and as to intensity. Such noise adds no variability to the work situation, but rather adds to its monotony, since the proportion of total stimulation which is nonvariable or monotonous is increased. It is tempting to maintain that within-session performance decrements occurring under such circumstances may be more adequately explained by a gradual reduction in arousal attributable to the low variability of the work situation than by an appeal to "psychological stress" (Jerison, 1959).

The present study lends no support to such a position. In all except the first (or training) session, whether or not noise was used, errors increased with time. A graphic representation of errors summated for each 6-minute period suggests both that the slopes of error curves for the Q, N, and L sessions are not significantly different and that the curve for the N session shows a considerably smoother rise. Incomplete records did not allow a more accurate check of these apparent trends.

Since there is no evidence that within-session performance decrement is affected by highly variable noise, it must be concluded tentatively that the arousal hypothesis does not account for the *decrease* in accuracy which is of most interest to investigators of vigilance tasks. Indeed, it seems that the increase in stimulus variability brought about by the noise acted uniformly throughout the session to reduce significantly the proportion of errors. Whether or not the error curve rises more uniformly under conditions of variability is a question for further investigation.

Alertness and Hypnotic Susceptibility

One way of interpreting the signaling of errors is as a measure of the extent to which the subject is critical of his performance. While peripheral to the question of the effect of noise on performance, the finding that those who signal errors most poorly are most susceptible to hypnosis deserves further mention. A critical attitude is a recognized deterrent to the induction of hypnosis. If a truck driver took such an attitude to the presence of a colonial house in the middle of the road (McFarland & Mosely, 1954, p. 124, Case 2) it is doubtful that an hallucination could become fully developed.

The magnitude of the relation between susceptibility to hypnosis and awareness of errors is indicated by the biserial correlation between high or low HI and the composite SI for the two sessions performed under Q conditions—$r_{bis} = .46$ ($p = < .01$). When the same statistic is computed, using a composite SI which also includes scores from the N sessions the resulting correlation is not significant ($r_{bis} = .28$). This suggests that susceptibility to hypnosis is most likely to be a performance relevant variable under highly monotonous work conditions.

Utility and Limitations of the Findings

The experimental monotonous task used in this study has been demonstrated to provide stable measures of errors (EI) and latency (LI), while further development might provide an equally adequate index of "awareness" of errors. These may reasonably be expected to predict performance in monotonous industrial jobs. However, there are certain considerations that suggest they be used and interpreted with caution.

The first of these is the finding from the present study that alertness appears not to be a unitary characteristic of the individual. A very careful criterion analysis must therefore precede attempts to validate such predictors, so that the one most relevant to desired performance may be used.

The second consideration has to do with the relation between the reaction of the individual to understimulating environments and his reaction to those which provide overstimulation. We have at present no indication of whether a relation exists between these two adverse reactions or, if it does, whether it is positive or negative. A positive relation would indicate the possibility of hiring workers whose performance would be relatively independent of conditions of environmental stimulation. If a negative relation exists it would suggest caution in hiring for monotonous jobs solely on the basis of this sort of predictor. In an extreme situation we might find our work force virtually helpless in the face of the overstimulation brought about by an emergency.

A third factor which should be investigated further is the question of adaptation to stimulus variability. Variability, by its very nature, cannot be augmented and prolonged indefinitely. Hence it will be important to know whether a certain level of variability continues to be associated with improved performance over an extended period of time.

Still another question to ask is whether predictive and concurrent validities for monotonous types of jobs and their predictors would be at all similar in magnitude. It may be that attrition or self-selection in monotonous jobs is an even more potent factor in range restriction than we have to deal with in most predictor variables. Hence we should be alert for the situation in which a barely significant relation, as determined on present employees, becomes a highly efficient predictor when used with applicants.

Finally, one must consider the question of motivation, since variations in motivation may play a critical part in the observed performance quite distinct from that played by the physical conditions of work. Considerable evidence has accumulated to show that unfavorable conditions of work may be compensated for within limits by increasing the incentive (Pepler, 1954). This suggests that unfavorable conditions may do more to inhibit a worker's *desire* to achieve greater production than his *ability* to do so. Similarly, variability introduced into the monotonous work environment may make it *possible* to perform more accurately without ensuring that more accurate work *will* result. In any short-range experiment there are bound to be idiosyncratic motivational effects which are atypical in terms of a routine work situation. For these several reasons a test in some sort of routine production situation will be necessary before the practical implications of this study can be assessed adequately.

Summary

The contradictory findings concerning the effects of noise on performance may be reconciled by a dimensional analysis of the characteristics of the noise and of the task involved, and an attack upon specific dimensions, based upon a theoretical rather than a strictly empirical approach. In terms of the "arousal hypothesis," noise which is low in "intelligibility" (or distraction value for the individual), while at the same time being high in variability, should enhance performance in a monotonous task.

The laboratory task which was devised to conform to the criteria established for a "monotonous" work situation consisted of 42-minute periods of handprinting pairs of letters in continuous sequence at predetermined, individually paced rates. Two groups of male subjects equated on performance in a first session, were exposed to noise of the specified type in counterbalanced order in

either a second or third similar session. A significant improvement in performance in terms of errors made in the task was associated with exposure to noise. Awareness of errors, as indicated by their signaling, was unchanged.

The similarity of phenomena associated with low stimulus environments and the hypnotic state suggested the obtaining of a standard measure of susceptibility to hypnosis from all subjects. Signaling of errors was significantly more accurate by those least susceptible to hypnosis, though the making of errors by subjects of low and high susceptibility was not significantly different.

In a fourth session a measure of latency of response to a series of unsystematically spaced visual signals from a low intensity light was recorded. The three measures of "alertness"—errors made, errors signaled, and latency—showed no significant intercorrelation, suggesting this characteristic is not unitary. The reliability of error and latency measures was sufficient to make them practicable in a personnel selection situation.

The implications of these findings are discussed in terms of their relation to "vigilance" tasks, and their applicability to personnel selection practice. The question of long-term adaptation to environmental variability is also considered.

REFERENCES

BERRIEN, F. K. The effects of noise. *Psychol. Bull.*, 1946, **43**, 141–161.

BEXTON, W. H., HERON, W., & SCOTT, T. H. Effects of decreased variation in the sensory environment. *Canad. J. Psychol.*, 1954, **8**, 70–76.

BROADBENT, D. E. Noise, paced performance and vigilance tasks. *Brit. J. Psychol.*, 1953, **44**, 295–303.

BROADBENT, D. E. *Perception and communication.* New York: Pergamon, 1958.

DEESE, J. Some problems in the theory of vigilance. *Psychol. Rev.*, 1955, **62**, 359–368.

FRIEDLANDER, J. W., & SARBIN, T. H. The depth of hypnosis. *J. abnorm. soc. Psychol.*, 1938, **33**, 453–475.

HEBB, D. O. Drives and the C.N.S. (conceptual nervous system). *Psychol. Rev.*, 1955, **62**, 243–253.

JERISON, H. J. Effects of noise on human performance. *J. appl. Psychol.*, 1959, **43**, 96–101.

MCBAIN, W. N. Use of psychological constructs for improving selection test validity. *Canad. J. Psychol.*, 1957, **11**, 164–170.

MCFARLAND, R. A., & MOSELEY, A. L. *Human factors in highway transport safety.* Boston: Harvard School of Public Health, 1954.

MELTON, A. W., & BRIGGS, G. E. Engineering psychology. *Annu. Rev. Psychol.*, 1960, **11**, 71–98.

MORGAN, J. J. B. The overcoming of distractions and other resistances. *Arch. Psychol., NY*, 1916, **5**, 1–84.

PEPLER, R. D. *The effect of climatic factors on the performance of skilled tasks by young European men living in the tropics.* (APU 199/53) Cambridge, England: Applied Psychology Research Unit, 1954.

SCOTT, T. H. Intellectual effects of perceptual isolation Unpublished doctoral dissertation, McGill University, 1954.

SCOTT, T. H., BEXTON, W. H., HERON, W., & DOANE, B. K. Cognitive effects of perceptual isolation. *Canad. J. Psychol.*, 1959, **13**, 200–209.

(Received September 19, 1960)

Journal of Applied Psychology
1961, Vol. 45, No. 5, 318–324

A VERIFICATION SCALE FOR THE STRONG VOCATIONAL INTEREST BLANK, MEN'S FORM

ROBERT W. FILBECK
University of Nebraska

AND

ROBERT CALLIS
University of Missouri

It is generally recognized that an occasional subject will respond to paper-pencil inventories in an invalid manner. Some subjects may fake their responses. Others will mark their answer sheets haphazardly or at random. Carelessness, improper motivation, and failure to comprehend inventory items or marking instructions are the more frequent reasons for invalid responding of a random or haphazard nature.

In view of the factors responsible for random responding being uncontrollable and undetectable during most test administrations, it seems advisable to devise, if possible, a means of identifying profiles derived from haphazardly marked answer sheets. This is particularly true of the Strong Vocational Interest Blank, Men's Form (SVIB), since this instrument is one of the most frequently used instruments in counseling and in research.

The problem posed here is not new. Strong (1943) determined, through die tossing, the characteristics of profiles derived from randomly marked SVIB answer sheets. Chance levels on each scale are denoted on the profile sheet as shaded areas. The vocational interests of dice then will fall within the shaded areas. The vocational interests of some real people also will fall in this area and the counselor is unable to determine with any reasonable degree of confidence whether a given profile is the result of chance answering or is actually a validly undifferentiated interest pattern.

Authors of other inventories similar to the SVIB have approached the problem of identifying random responding through construction of special scales (Callis, 1948; Hathaway & McKinley, 1945; Kuder, 1956). Such scales, termed "validity" or "verification" scales (V scales), were constructed by scoring for item responses chosen rarely in real testing and significantly more often by chance. (The converse was true for the Kuder V scale.)

The Problem

The purpose of this investigation was to construct and validate a V scale for the SVIB. The V scale construction procedures approximated those employed by Kuder, Hathaway and McKinley, and Callis in developing "rare response" scales for other inventories.

Since the rationale underlying a rare response scale defines deviant scores in the chance direction as being indicative of either random responding or of highly unusual interests, an attempt was made to construct a scale that would provide a range of scores that would be obtained infrequently either in the student population or from random responding. This "genuine but unusual" level would fall at an intermediate position to the typical and chance levels on the V scale. It was hypothesized that the intermediate level would be provided by including items in the scale that were rarely selected by college students, but of a lesser degree of rarity than commonly required. For this reason, the criterion for rarity was a difference between obtained and chance response frequencies significant at the .10 level rather than the more stringent criteria of the .05 and .01 levels.

Procedure

The construction sample was composed of 192 freshmen enrolled in the University of Missouri during the 1955–56 school year. The total sample was divided into eight division-by-sex subsamples as follows:

MALES	FEMALES
Arts and Science	Arts and Science
Agriculture	Agriculture
Education	Education
Engineering	Nursing

Each subsample, except males in Education, contained 25 subjects. Only 17 subjects were available for inclusion in the Education males subsample.

The SVIB had originally been administered to a group of approximately 400 freshmen for purposes of a larger study, and the SVIB answer sheets of the construction sample were selected by random means from this population.

The first step in isolating rare item-responses consisted of contrasting the response frequency for the total construction sample with the frequency expected from chance (33%, since there are three item-responses for each item). The index used was the phi coefficient (Jurgenson, 1947). The criterion for rarity was a phi significant at the .10 level ($\Phi = .127$), with the obtained frequency being less than the chance frequency.

Next, the response frequency for each item response was contrasted with chance expectancy for each division-by-sex subsample. The criterion for acceptance here was chance frequency exceeding the obtained frequency by 10%.

Results

These procedures yielded a V scale of 103 item-responses, all unit weighted in the positive direction, and distributed among 97 items. The SVIB item numbers and item responses scored on the V scale key are listed in Table 1.

Validity

Initial validation of the SVIB V scale was a test of the ability of the scale to discriminate between genuine and chance marked answer sheets. Since, with large numbers, relatively small scale distances between means will be statistically significant, a difference-between-means criterion for validity was not used. Instead, percentage of overlap of the two criterion distributions was considered a proper index of validity. If the two distribution means were clearly separated and overlap was so slight as to permit rather unambiguous dichotomization, the scale would be considered valid for the purpose stated here.

TABLE 1

SVIB Items and Response Categories Scored on the SVIB V Scale

15.	L	204.	D	261.	D	346.	3
34.	I	205.	D	262.	D	356.	3
36.	L	206.	D	264.	D	357.	1
38.	L	207.	D	265.	D	358.	3
54.	L	208.	D	268.	L	359.	3
68.	L	227.	D	270.	L	360.	1
75.	L	228.	D	271.	D	363.	no
91.	L	234.	D	272.	L	367.	no
97.	L	236.	I, D	273.	L	370.	?, no
98.	L	237.	L	275.	L	371.	no
137.	D	238.	L	276.	D	372.	no
140.	D	239.	L	277.	L	375.	no
141.	I, D	241.	L	279.	D	376.	no
154.	I, D	242.	D	280.	D	381.	no
155.	D	245.	D	284.	I	382.	no
166.	D	250.	D	298.	1	383.	no
173.	D	251.	L	315.	1	384.	no
174.	D	252.	I, D	317.	3	385.	no
175.	L	253.	L	325.	1	386.	no
176.	L	254.	L	327.	2	387.	no
177.	L	255.	D	329.	3	388.	?, no
182.	D	258.	L	337.	3	390.	yes
183.	D	259.	D	340.	2	394.	no
196.	D	260.	L	342.	3	398.	no
						400.	no

The cross-validation sample was composed of 908 subjects; 409 male and 249 female freshmen from the University of Missouri, 100 male clients selected at random from the files of the University of Missouri Testing and Counseling Service, 100 male freshmen from the University of Minnesota,[1] and 50 males of upper class and graduate status from Central Missouri State College. The V scale mean and *SD* for the total cross-validation sample were 8.55 and 4.73, respectively. The mean and *SD* of the theoretical chance distribution were computed to be 34.4 and 4.6, respectively.[2] The two distributions are depicted in Figure 1.

Examination of Figure 1 reveals the validity criterion to be amply satisfied. The two distributions are clearly differentiated, more than five *SD*s separating the two means.

While it is obviously impossible to establish a cutting score which will permit perfect dichotomizing of V scores into real or chance categories, it is possible to sort answer sheets into real or chance groups with about 99% accuracy if the cutting score is set at V score equal 23. Just less than 1 in 100 chance-an-

[1] The writers wish to thank Ralph Berdie for his cooperation in furnishing SVIB records from the University of Minnesota.

[2] Statistics of the theoretical chance distribution were computed from the following formulas:

$$\bar{x} = \frac{N_1 + 2N_2}{3}$$

$$SD = \sqrt{N_1PQ + N_2PQ}$$

where N_1 = number of items contributing one item response

N_2 = number of items contributing two item responses

P = probability of chance marking of V scale item-response for each item

Q = probability of chance marking of non-V scale item-response for each item.

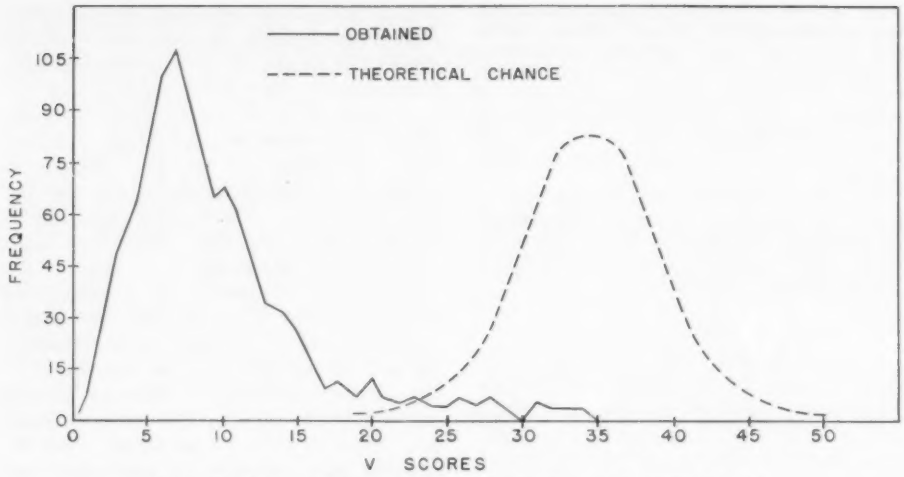

FIG. 1. Obtained and theoretical chance distributions on the SVIB V scale.

swered inventories will have V scores of less than 23, while 1.29% of the assumed genuinely answered inventories score at this value or higher.

Sex and Academic Group Differences on the V Scale

Although the procedures followed in the construction of the V scale held to a minimum the likelihood of obtaining V scale differences among student groups classified as to academic affiliation or sex, it was desirable to determine on a cross-validation sample if this, in fact, was accomplished. For this test the distributions of the four male and four female validation subsamples from the University of Missouri were contrasted. The means and *SD*s of these subsamples are presented in Table 2.

The data in Table 2 indicate the construction procedures to have been successful in providing a scale of item responses which are rarely selected by university students, regardless of sex or curricular affiliation. However, even the small differences found here are too great to be charged to chance sampling effects. Analysis of variance techniques yielded an *F* ratio of 2.21, significant at between the .04 and .03 levels.

The finding of a significant value of *F* in this instance does not require any substantial modification of interpretation of a client's V score because of his curricular affiliation however. The differences between groups are too small, two raw scores separating the means of the two extreme groups, to be interpretable in the practical sense.

Institutional Differences on the V Scale

To test the possibility that populations in institutions of higher education other than

TABLE 2

V SCALE MEANS AND STANDARD DEVIATIONS OF EIGHT UNIVERSITY OF MISSOURI CROSS-VALIDATION SAMPLES

Sample	*n*	*M*[a]	*SD*
Males			
Arts and Sciences	159	9.5	5.3
Agriculture	113	8.3	4.7
Engineering	137	7.6	4.4
Counseling Clients	100	9.6	5.3
Females			
Arts and Sciences	117	9.3	5.0
Agriculture	14	9.7	3.8
Education	84	8.9	5.2
Nursing	34	8.8	4.3
Total	758	8.9	5.0

[a] An analysis of variance test for significance of the differences among these means indicated only a slight tendency to significance ($F = 2.21$; $.04 < p > .03$).

the University of Missouri will score differently on the V scale, the V scale distributions of the male samples from the three aforementioned institutions were contrasted. The means and *SD*s of the three samples appear in Table 3. Determination of the likelihood of the sample differences arising by chance effects was by analysis of variance techniques. An *F* ratio of 2.76, not significant at the .05 level, was obtained; offering no contrary evidence to a hypothesis of random sampling from populations of equal means.

Reliability

Internal consistency of the V scale was estimated from the total University of Missouri sample, using Kuder-Richardson Formula No. 3 (Adkins, 1947). An *r* of .66 was obtained, which may be considered an underestimate since the assumption of equal item difficulty was not met. The desired degree of internal consistency is difficult to specify. There are no data available which would suggest that such a V scale should have high or low internal consistency.

D. Brantley (1958 unpublished) readministered the SVIB with standard instructions to a group of 25 males, allowing a 2-week interval between administrations. He obtained a V scale test-retest *r* of .80.

Relationship of V Scale to Other SVIB Scales

Using the 100-male freshmen constituting the University of Minnesota sample, Pearson product-moment correlations were computed between the V scale and each of the scales on the SVIB profile. In the main, the *r*'s were low, less than half departing significantly from zero. The *r*'s which were significant ($p < .05$ and $p < .01$) are listed in Table 4.

TABLE 3

V Scale Means and Standard Deviations of Male Samples from Three Institutions of Higher Education

Institution	*n*	*M*[a]	*SD*
Central Missouri State College	50	8.3	3.44
University of Minnesota	100	8.5	4.59
University of Missouri	509	8.3	4.76

[a] An analysis of variance test for significance of the differences among these means indicated no significant differences ($F = 2.76$; $p > .05$).

TABLE 4

Significant Correlations between SVIB Profile Scales and the SVIB V Scale

Scale	*r*
Artist	.24*
Veterinarian	.21*
Farmer	.20*
Aviator	.21*
Printer	.21*
Forest Service Man	.22*
YMCA Physical Director	.28**
Personnel Director	.28**
Public Administrator	.40**
YMCA Secretary	.26**
Social Science High School Teacher	.37**
City School Superintendent	.40**
Social Worker	.28**
Musician	.47**
CPA	.26**
Senior CPA	.26**
Advertising Man	.20*
Lawyer	.23*
Author Journalist	.30**
Occupational Level	.20*

* $p < .05$.
** $p < .01$.

Noting the tendency for higher V scores to accompany interests in the social welfare area, 19 profiles having a Group V primary interest pattern were contrasted to those not so characterized. The "Group V" profiles had a mean V score of 7.1, the contrasted group had a mean V score of 8.8. Failure to obtain a difference in the expected direction in this test would seem to indicate that no modification of V scale interpretation need be made on account of measured interests.

V Scale Norms

Standard score equivalents of SVIB V scores are given in Table 5. Table 5 was derived from the V scores of the total cross-validation sample of 908 subjects.

Because of marked skewness of the student distribution (see Figure 1) scaling based on normal curve procedures was inappropriate, so the *T* score equivalents in Table 5 were derived from an area transformation (Guilford, 1950).

TABLE 5

NORMS FOR THE SVIB V SCALE

Raw Score	*T* Score[a]	Raw Score	*T* Score[a]
0	0	17	65
1	21	18	66
2	27	19	67
3	31	20	68
4	36	21	70
5	40	22	70
6	43	23	71
7	46	24	72
8	49	25	72
9	51	26	73
10	53	27	74
11	55	28	74
12	57	29	75
13	59	30	76
14	61	31	76
15	62	32	79
16	64	33	81

[a] Standard scores with a mean of 50 and a standard deviation of 10 obtained by an area transformation (Guilford, 1950).

Interpretation of SVIB V Scores

In the statement of the problem it was pointed out that use of the V scale should permit classification of SVIB answer sheets into three categories: genuine and typical, genuine but deviant, and chance or random response. From Figure 1, it is apparent that the genuine/typical and the chance categories are clearly differentiated. Also apparent is the existence of a range of scores near the cross-over point of the two distributions that includes relatively few cases from either distribution.

Establishing cutting scores to define the limits of each category, in the absence of empirical evidence, has been approached arbitrarily, relying on the experiences of authors of other scales.

Earlier, it was demonstrated that the likelihood of mistakenly accepting a chance-answered inventory could be held to less than 1 in 100 at the expense of rejecting approximately 1.3% of the presumably genuinely answered inventories as being responded to at random. Ordinarily, rejection of slightly more than 1% would be deemed a reasonable "price" to pay for objective assurance of valid testing. However, experience with the *F* scale of the Minnesota Multiphasic Personality Inventory would indicate that even this small percentage may be too great. Hathaway and McKinley (1945) point out that high scores ($T = 65$–80) indicate unusual behavioral characteristics but do not invalidate the testing. They do not consider the testing questionable at less than $T = 80$.

Mindful of this experience, the suggested cutting score between genuine and random responding on the SVIB V scale is set at 28 ($T = 74$). This will hold the likelihood of including a chance answered inventory to less than 9 in 100, a reasonable confidence level for the purposes specified here.

The lower boundary of the intermediate or "valid but highly unusual" level is more difficult to specify rationally or empirically. Hathaway and McKinley use the $T = 65$ score as a dividing point on the MMPI *F* scale. In the absence of empirical data, a similar dividing line is suggested here, thus defining as "highly unusual interests" SVIB profiles having V scores 65 through 74 (raw V scores 17 through 28).

Suggesting the boundaries of unusual interests does not, however, define the nature nor the extent of "unusualness." Carnes [3] has provided some preliminary information in this respect in reporting on the V scale distribution obtained from the SVIBs of 40 hospitalized patients in a Veterans Administration psychiatric hospital. He had personally administered the inventories under highly individualized conditions, thus giving subjective assurance of valid test taking. The mean patient V score was $T = 70$ (raw score 21.8) and the patient sample *SD* was 14. The information Carnes provides suggests that in many individuals, high V scores reflect strong conflicting drives which tend to be disorganizing to the total personality and are more determinative of behavior than are interests. In predicting the activities an individual with a high V score would find satisfying, the counselor would probably place minimal dependence on the SVIB profile as such, and would resort to other, more clinical, indications of drives and motives.

[3] C. D. Carnes, personal communication, 1953.

Effects of Faking on V Scores

Although the SVIB V scale was not designed as a device to detect faking, one investigation has studied the effects of deliberate falsification of interests on V scores.

D. Brantley (1958 unpublished) administered the SVIB to a group of 25 men enrolled at Central Missouri State College. The first adminstration was according to standard instructions, the second to fake. The faking instructions directed subjects to respond to the SVIB as though applying for a highly remunerative but intrinsically uninteresting job, to be selected from the occupations listed on the SVIB booklet. The V scores of the faked answer sheets showed a mean increase of six raw score points over the sincere answer sheets. Although this increase is significant ($t = 2.56$, $p < .05$), the mean faking score does not approach the unusual interest range. Faking, therefore, does not appear to be an hypothesis to be attached to higher V scale levels.

Summary

A "verification scale" (V scale) has been constructed for the Strong Vocational Interest Blank, Men's Form. Item analysis of the SVIB, based on the responses of 192 male and female subjects, isolated 103 item-responses which were rarely selected by students. "Rarity" was determined by contrasting the obtained frequency of response with that expected from a chance answering of the SVIB. The criteria for rarity were differences significant at the .10 level, for the total group, and 10% difference between obtained and chance frequencies for each major curricular and sex group.

Validation studies to date, utilizing a total cross-validation sample of 908 subjects, yield the following information:

1. Genuine and chance response profiles are well differentiated by the SVIB V scale.

2. The differences among the mean V scores of curricular groups and between sexes are minor, although there is an indication that males in engineering tend to obtain slightly lower V scores.

3. No significant differences were found among the mean V scores of three male samples from different institutions of higher education.

4. V scores in the range $T = 65$–75 suggest personality deviation of a nature requiring some modification of SVIB profile interpretation.

The reliabilities of the V scale were estimated to be .66 by means of K-R Formula #3 and .80 by test-retest technique when the interval between testings was two weeks.

Coefficients of correlation between the V scale and the SVIB profile scales show a significant value of r in 20 instances, although the r's tend to be small. It was determined, however, that no differences in V scale interpretation need be made because of measured vocational interests.

On the basis of present information, counselors may use the V scale to establish three general levels of scores:

1. T scores of 0–64 represent a range of scores which typify answer sheets marked with attention and understanding and presumably on the basis of "normal" interests. The interpretation of SVIB profiles with V scores in this range may proceed rather straightforwardly.

2. T scores of 65–74 represent a range of scores which indicate an increasing likelihood of random responding or of deviant personal chracteristics. The interpretation of profiles with V scores in this range should be more tentative, exploring first the possibilities of random responding and then of strong conflicting drives which will tend to negate or modify measured interests as predictors of future behavior.

3. T scores of 75 and above are strongly indicative of random responding or of extremely unusual interests. In these extreme cases the counselor would likely readminister the SVIB with attention being given to client understanding of inventory content and marking instructions. V scores in this high range, if shown to be the result of attentive and comprehending responding, will probably invalidate use of the SVIB as a measure of interests.

At the present stage of development it is appropriate to recommend that the V scale be included among the scales currently scored on

the SVIB. Its usefulness in establishing response validity is sufficiently established and research on the implication of atypical V scores will be expedited.

REFERENCES

ADKINS, DOROTHY C. *Construction and analysis of achievement tests.* Washington, D. C.: United States Government Printing Office, 1947.

CALLIS, R. Change in teacher-pupil attitudes related to training and experience. Unpublished doctoral dissertation, University of Minnesota, 1948.

GUILFORD, J. P. Fundamental statistics in psychology and education. (2nd ed.) New York: McGraw-Hill, 1950.

HATHAWAY, S. R., & MCKINLEY, J. C. Manual for the MMPI. New York: Psychological Corporation, 1945.

JURGENSON, C. E. Tables for determining phi coefficients. *Psychometrika,* 1947, **36**, 17–29.

KUDER, G. F. *Manual for the Kuder Preference Record, Form C.* Chicago: Science Research Associates, 1956.

STRONG, E. K., JR. Interests of men and women. Stanford: Stanford Univer. Press, 1943.

(Received September 23, 1960)

Journal of Applied Psychology
1961, Vol. 45, No. 5, 325-329

PERSONALITY AND PRODUCT USE

W. T. TUCKER AND JOHN J. PAINTER

University of Texas

Perhaps no subject in marketing has received greater attention in the past few years than the relationship between personality and purchasing behavior. All of the furor over motivation research is clearly predicated on the premise that such a relationship exists, although some reporters seem to assume that all persons are, at base, alike. Yet even here, the factors referred to as common to all persons are most often those which personality studies have shown to be variables rather than constants. For instance, the importance of fear of the father image, which is reputed to militate against the use of banking services, must be conceived of as varying with some personality characteristic such as ego strength or emotional maturity if it is not to influence all persons in a highly similar way.

Talk about the importance of personality as a marketing variable has become common at advertising clubs and at marketing association meetings. The recent book by Pierre Martineau (1957) contains a chapter entitled "An Automobile for Every Personality." Charles Cannell (Ferber & Wales, 1958) says: "It may be that the determination of airplane travel has something to do with basic personality characteristics such as personal feelings of security or insecurity" (p. 10). And Ernest Dichter (Ferber & Wales) says confidently: "What we are searching for are psychological and personality elements which may have a dynamic effect on consumers' attitudes toward a product" (p. 26). Newman (1957) views personality as one of the major factors determining marketing behavior.

In the light of such points of view it may seem surprising that few efforts have been made to demonstrate that personality characteristics actually do influence product use. But the dearth of evidence on this point can be explained in part by supposition. First, the concept of personality itself has not been very clearly formulated. Second, the instruments available for the ready classification of personality types are few and generally suspect. Third, most self-respecting psychologists are apparently convinced that marketing behavior, pervasive as it may be, is of interest for commercial purposes only. Fourth, marketers probably have little understanding of the need for experimental evidence of their assumptions.

Yet it would seem that there is much to be learned about both personality and a large segment of human behavior by such studies. Scott's (1957) study of motion picture preferences is perhaps of less interest to the movie producer than it is to the individual who wants a clearer understanding of the personality factors isolated by the Minnesota Multiphasic Index. That these factors are less than completely clear is indicated by Scott's inability to provide a rationale for all of the significant correlations. And Eysenck's (Eysenck, Tarrant, Woolf, & England, 1960) recent findings that rigidity and extraversion relate to the number of cigarets smoked by an individual may be as important to the understanding of those characteristics as they are as a possible explanation of lung cancer in heavy smokers.

The present study was undertaken to test the hypothesis that marketing behavior is related to personality traits. At the same time, it was expected that the location of significant relationships would throw additional light on the meaning of personality characteristics studied.

METHOD

The Gordon Personal Profile was administered to 133 students of marketing along with a so-called Sales and Marketing Personality Index which included questions on the use of headache remedies, cigarets, chewing gum, deodorants, mouthwash, and other items commonly purchased by college students. Blind questions were interspersed to give the index the appearance of a personality or interest test. Results were then compared to determine the difference in personality trait scores for groups that professed to different rates of product use or interest. That the subjects accepted the index was indicated by the large number of students who asked after

completing the forms if they could find out whether they would make good salesmen, advertisers, etc.

Subjects

The subjects were all students of the first course in marketing at the University of Texas. The great majority were juniors; a few were in the last semester of their sophomore year, and others were in the beginning of their senior year. Since the Gordon Personal Profile has different norms for male and female students, and, since the frequency of use of a number of products was clearly related to sex, the 31 responses by females are not included in this report. Also, one subject was eliminated because he failed to fill out the Gordon Personal Profile completely. While this group of subjects can hardly be characterized as representative of even such a limited universe as college juniors, for purposes of this study their only necessary characteristic was that of providing a diverse group of scores on the Gordon Personal Profile and reasonable diversity in response to questions about products.

Test Materials

The Gordon Personal Profile was selected as the personality test to use since it measures four characteristics which seem intuitively meaningful as components of the "normal" personality and since it is based on college student norms. The profile rates persons on the variables of ascendency, responsibility, emotional stability, and sociability.

The form used to determine use of products or other marketing characteristics included nine questions relevant to the experiments and seven blind questions. Most of the experimental questions referred to frequency of use of a particular product, as in the following:

> How frequently do you experience a headache that requires a headache remedy (aspirin, Bufferin, Anacin, etc.)
>
> a. Never
> b. Once or twice a year
> c. About once a month
> d. More often than once a month, but less than once a week
> e. Once a week or more

Questions of this sort were asked about the use of headache remedies, vitamins, chewing gum, tobacco, mouthwash, alcoholic beverages, and deodorants. Two other questions related to the readiness with which the individual accepted new styles or fashions and preference in automobiles.

Blind questions were rather similar to those asked on interest tests:

> Which of the following positions in an organization would you prefer to hold?
>
> a. Secretary-Treasurer
> b. Program chairman
> c. President
> d. Membership chairman
> e. Ordinary member, no office

The list of 16 questions was pretested in order to insure their clearness and to make sure that multiple choice answers would elicit a reasonable spread of response. As a result of this pretesting, multiple choice answers were altered to fit the normal variations in frequency of use of various products. For instance, the most frequent use of headache remedies indicated by answers was "once a week or more," while the most frequent use of deodorants was described as "more than once a day," since the pretest demonstrated these to be common frequencies for heavy users.

Procedure

Subjects filled out both forms at a single sitting of about 20 minutes, answering the Gordon Personal Profile first and the Sales and Marketing Personality Index second. While subjects were asked to fill in their sex, age, marital status, and year in school on the Gordon Personal Profile, names were not taken in order to encourage the greatest frankness in response. Each pair of tests handed out was numbered in advance.

Instructions to Subjects

Students in each class tested were given the following instructions:

> As you all know, one of the difficult problems in business is the determination of an individual's interests, or what kind of job he can do best. Attempts to solve such problems have led to the development of a number of written tests—some of which take an hour or more to complete. You have in front of you two rather new tests that try to accomplish this for certain marketing jobs in just a few minutes. We know that one of these is moderately successful. We are interested in whether scores on the other are different or much the same. Do these tests really measure the same things?
>
> To determine this, we need your help. We are not interested in your score as an individual but in the relationship of your score on one test with your score on the other.
>
> For that reason, we do not want your name on the paper; we merely want you to answer the questions honestly and conscientiously. Instructions for each test appear at the top of the test.
>
> First, make sure the red number on each of your tests is the same. Then fill in your age, sex, marital status, and year in school on Test #1, the Gordon Personal Profile. Then read the instructions on that test and answer the questions. When you are finished, go directly to the second test, read the instructions, then answer the questions.
>
> You will find that on both tests there are some questions where none of the answers seem just right for you. Just pick the one that seems closest

TABLE 1

MEAN RESPONSIBILITY SCORES FOR GROUPS WITH DIFFERENT PATTERNS OF USE OF MOUTHWASH

Response	Mean Score	Number of Cases
Never use mouthwash	7.26	31
Quite infrequently	5.00	40
Once or twice a week	4.50	16
Once a day	3.90	10
More than once a day	6.50	4

and do not worry about exact wording. Remember to read the test instructions carefully, since you have to answer each test in a somewhat different way.

Analytic Method

Results were analyzed by comparing the difference in mean scores on one personality characteristic for groups with different product use patterns.

Table 1 shows the mean scores on responsibility for groups which answered the mouthwash question in each of the possible ways.

While responsibility seems to be inversely related to frequency of use of mouthwash, despite relatively high scores for the four persons who use mouthwash more than once a day, the number of cases in some of the cells is too small for analysis of variance to show a significant relationship. For this reason, the last four groups were combined and compared using the t test with those who reported never using mouthwash. The resulting t of 2.12 is significant at the .05 level. The F test for homogeneity of variance was not significant.

This same method of analysis was used for each of the products on each of the personality characteristics, with the point for division into two groups being determined on the basis of scores and the number of subjects remaining in each of the groups.

It is entirely proper to question whether five-point scales of the sort used here should be dichotomized *after* observing the means of each of the categories. Such dichotomization obviously makes it possible to maximize the number of "significant" relationships. Where possible, it should therefore be avoided and some independent method should be used for dichotomization.

Since there is no apparent rationale for predicting relationships between personality characteristics and product use, it seems foolhardy to develop a purely arbitrary dichotomization method in the present case. Such a method could easily minimize relationships if it were only extremes of product use that related to personality measures and the cutting point closest to the median were arbitrarily used, for instance. It happens that dichotomizing the data shown in Table 1 by combining the top two categories and the bottom three does not lead to a significant t. The resulting quandary is more philosophical than statistical. It seems to the authors that refusing to locate the cutting point that leads to statistically significant differences is the more serious error when dealing with the kind of problem discussed here.

RESULTS AND DISCUSSION

A total of 36 comparisons (9 product categories × 4 personality characteristics) included 13 significant relationships. As might be expected, some products were associated with no personality trait; others were associated with one or more; and one product, vitamins, was associated with all four of the personality traits.

Table 2 shows those relationships indicating the significance level. In addition it shows correlation ratios in parentheses to indicate the approximate strength of the relationships.

The results clearly indicate that there is a relationship between product use and personality traits. This relationship apparently may include both frequency of use of a particular product and preference among different brands of a single product, since preference in automobiles is significantly related to scores on responsibility. At the same time, some products are used frequently or infrequently without relationship to any of the personality traits tested. Each personality trait seems to bear a relationship to the use of some products, each of the four traits scored by the Gordon Personal Profile relating to the use of at least two of the products considered in the present experiment.

It should be pointed out, however, that the relationships located between product use and personality are not particularly strong, certainly less strong than popular marketing concepts of the day suggest.

An obvious corollary to the conclusion that personality traits and product use are related is that the Gordon Personal Profile does isolate personality traits related to behavioral differences. Further, an examination of the pattern of significant relationships shown in Table 2 is persuasive that the four traits, ascendency, responsibility, emotional stability, and sociability, have considerable independence. The manual for the test indicates that the intercorrelations are generally low except for those between ascendency and sociability (.43) and between emotional stability and re-

TABLE 2

SIGNIFICANT PERSONALITY TRAITS IN THE USE OR PREFERENCE FOR SOME CONSUMER PRODUCTS

	Ascendency	Responsibility	Emotional Stability	Sociability
Headache remedies	−.05 (.464)[c]	—	−.05 (.320)[c]	—
Acceptance of new fashions	.01 (.331)[c]	—	—	.01 (.566)[c]
Vitamins	−.05 (.332)[c]	−.01 (.297)[c]	−.01 (.091)[c]	−.05 (.272)[c]
Cigarettes	—	—	—	—
Mouthwash	—	−.05 (.224)[c]	—	—
Alcoholic drinks	—	−.01 (.362)[c]	—	—
Deodorants	—	—	—	—
Automobiles[a]	—	.01 (.281)[c]	—	—
Chewing gum[b]	—	.05 (.295)[c]	.01 (.331)[c]	—

Note.—In all cases except for the last two products, the sign indicates the nature of the relationship. High ascendancy is related to infrequent use of headache remedies, for instance, but with rapid acceptance of new fashions.

[a] Subjects who preferred the more popular makes of car such as Buick, Dodge, Mercury, Ford, Chevrolet, and Plymouth rated higher on the responsibility scale than those who stated a preference for such sports cars as the Corvette or Thunderbird.

[b] While there is no significant difference in personality trait scores and the amount of gum chewed, those who chew gum *only when offered it by someone else*, are significantly lower than others in responsibility and emotional stability.

[c] Correlation ratios.

sponsibility (.46). Those correlations were considerably higher in the present experiment as shown in Table 3. The remaining correlations are quite low. It must be concluded that the Gordon Personal Profile does not measure four independent characteristics but two independent sets of related characteristics. However, it seems that one of a set of related characteristics can still prove to have enough relative independence to be conceptually valuable.

Most of the significant relationships between product use and character traits located are intuitively acceptable. One would expect that high ascendency and high sociability would be related to the rapid acceptance of new fashions, especially since ascendency is described largely as social leadership. On the other hand there seems to be no particular reason for expecting all personality characteristics to be associated with the frequency of use of vitamins, unless one conceives that personality traits are most likely to affect behavior that society neither rewards nor punishes.

TABLE 3

INTERCORRELATIONS AMONG PERSONALITY TRAITS

Traits	Ascendency	Responsibility	Emotional Stability
Responsibility	.058		
Emotional Stability	.035	.695	
Sociability	.708	.035	.086

The results cast some possible light on the nature of responsibility as a character trait. It is related to avoidance of vitamins and mouthwash, preference for popular cars and moderate drinking or abstinence. Since these are all modal characteristics of the group being tested, a reasonably strong case might be made for the fact that responsibility is closely related to the acceptance of group norms.

A comparison of the present results with those of Eysenck (1960) suggests that sociability on the Gordon Personal Profile is considerably different from extroversion, with which it might seem related. Eysenck's results showed a strong, significant correlation between extroversion and heavy cigarette smoking, while the present experiment did not even hint at such a relationship between sociability

and heavy smoking. It is possible that the difference in age (Eysenck's subjects were considerably older) or difference in nationality (Eysenck's subjects were British) might explain this apparent contradiction.

Summary

The answers to the Gordon Personal Profile and a disguised product use questionnaire by 101 college of business students demonstrate that personality traits are often related to product use. Thirteen of a possible 36 such relations were significant at the .05 level or above.

A corollary conclusion is that the Gordon Personal Profile distinguishes personality traits related to behavioral differences, although the four traits are not "independent."

References

Eysenck, H. J., Tarrant, Mollie, Woolf, Myra, & England, L. Smoking and personality. *Brit. med. J.*, 1960, **5184**, 1456–1460.

Ferber, R., & Wales, H. *Motivation and marketing behavior*. Homewood, Ill.: Irwin, 1958.

Martineau, P. *Motivation in advertising*. New York: McGraw-Hill, 1957.

Newman, J. W. *Motivation research and marketing management*. Boston: Harvard Univer. Press, 1957.

Scott, E. M. Personality and movie preference. *Psychol. Rep.*, 1957, **3**, 17–18.

(Received October 10, 1960)

Journal of Applied Psychology
1961, Vol. 45, No. 5, 330–337

FACTOR ANALYTIC DEFINITIONS OF VOCATIONAL MOTIVATION

JOHN O. CRITES [1]

State University of Iowa

Reflecting the increasing emphasis during the past decade upon the role of dynamic factors in occupational choice and vocational adjustment (Ginzberg, Ginsburg, Axelrad, & Herma, 1951; Roe, 1956), recent research on the measurement of vocational phenomena deals largely with the construction of instruments designed to assess such motivational variables as preferences for different occupational activities and situations, evaluations of the desirability of various job components, and judgments about the relative appeal of tangible and intangible work incentives. Among the new measures of vocational motivation are the Occupational Attitude Rating Scales (OARS) developed by Hammond (1954, 1956, 1959), the Work Satisfaction Questionnaire (WSQ) devised by Astin (1958), and the Job Incentive Rankings (JIR) constructed by Bendig and Stillman (1958). Based upon factor and cluster analyses of item ratings and rankings, these self-report questionnaires yield scores on 11 vocational motivation scales: Materialistic, Competitive, Technical, and Humanitarian (OARS); Managerial-Aggressive, Status-Need, Organization-Need, and Working Conditions [2] (WSQ); and, Achievement-Need, Intrinsic Job Interest, and Job Autonomy (JIR). Within the three inventories, due to their factorial composition, scores on one scale are essentially independent of those on another, the highest correlation being .28 for the Materialistic and Competitive scales of the OARS (Hammond, 1956).

The intercorrelations of the scales from the various inventories are unknown, but it seems reasonable that they are related and define similar variables. Both Hammond (1956) and Astin (1958) found differences between college major groups on the OARS and WSQ which indicate possible relationships between the scales of these questionnaires. For example, the Materialistic (OARS) and Managerial-Aggressive (WSQ) scales differentiate business students from those in other curricula, and the Technical (OARS) and Organization-Need (WSQ) scales distinguish between science and nonscience majors. Bendig and Stillman (1958) report no group comparisons, but the item content of the JIR scales suggests that they measure attitudes much like those assessed by the OARS and WSQ. To illustrate, the Materialistic (OARS) and Job Autonomy (JIR) scales contain almost identical items ("economic security" and "good job security"), and the Managerial-Aggressive (WSQ) and Achievement-Need (JIR) scales express quite similar preferences and values ("A job where I do *not* work under instructions" and "Freedom to assume responsibility"). There is some evidence, then, that the scales of the newly developed vocational motivation inventories are not independent and that they cluster into distinguishable groups.

The purpose of the present study was to test for the existence of interrelationships among the OARS, WSQ, and JIR scales and to identify the groups which the scales form. More specifically, the objective was to reduce the number of scales to a more parsimonious set and to provide factor analytic definitions of the dimensions of vocational motivation. The anticipated outcomes of the analysis were a clarification of the structure of vocational motives, an evaluation of the composition of the inventory scales, and a basis for the integration of theoretical constructs.

PROCEDURE

Measuring Instruments. The Occupational Attitude Rating Scales contain 40 items, 10 for each

[1] The author wishes to express his appreciation to Harold P. Bechtoldt for his generous assistance in processing and analyzing the data.

[2] Although Astin (1958) assigns no name to the item cluster which constitutes this scale, for convenience in discussion it is called "Working Conditions."

of four scales, which a subject rates on a five-point scale. The scale extends from "much liked," through "indifferent," to "much disliked." The scores for the scales are the unweighted totals of keyed items. Internal consistency estimates for the scales range from .66 (Materialistic) to .80 (Humanitarian) (Hammond, 1956). The Work Satisfaction Questionnaire has 19 statements about different kinds of work which a subject rates on a seven-point scale of desirability. Scale scores are the sums of the ratings. Astin (1958) reports no reliability data for the scales of the questionnaire. The Job Incentive Rankings are made with 8 verbal descriptions of job incentives, which the subject orders according to their importance to him in choosing an occupation. Scores are obtained by subtracting the ranks of pairs of job incentives which define the three scales. Bendig and Stillman (1958) present no findings on the reliability of the rankings.

Subjects. The sample consisted of 300 subjects in a large undergraduate, introductory course in psychology. All subjects had at least sophomore class standing. They took the OARS, WSQ, and JIR voluntarily as part of the course and received additional credit for participating in the test administration. In the sample there were equal numbers of males and females. Because preliminary analysis of the data indicated possible sex differences in intertest relationships, the male and female variance-covariance matrices were compared for heterogeneity. The findings ($t = .72$, $df = 131$, $p > .05$) supported the homogeneity of the matrices, however, and males and females were combined for the factor analysis proper, which was based upon the entire sample of 300 subjects.

Method. The 11 variables for the OARS, WSQ, and JIR were intercorrelated by the usual product-moment techniques, and the resulting matrix of 55 coefficients was factor analyzed by the complete centroid method, after the hypothesis of independence among the inventory scales was rejected (R. Bargmann, 1957 unpublished). Four factors were iterated through two cycles by Bargmann's procedure, and Lawley's (1951) maximum likelihood test of the independence of residual covariances was applied to determine whether additional common factors remained. Since the test was significant ($\chi^2 = 23.49$, $df = 10$, $p < .01$), one more iteration was made, and nonsignificant residuals which ranged from $-.048$ to .046 were obtained. With analytic (oblimax) and graphical methods, the resulting five factors were then rotated to an orthogonal simple structure solution.

Results

Table 1 presents the intercorrelation matrix for the OARS, WSQ, and JIR scales, as well as the means and standard deviations for each variable. As compared with total possible score ranges, and in ratio to the means, the standard deviations for the scales indicate adequate variability for the purposes of correlational and factorial analyses. From inspections of the score distributions for all

TABLE 1

Intercorrelations of Occupational Attitude Rating Scales (OARS), Work Satisfaction Questionnaire (WSQ), and Job Incentive Rankings (JIR) Variables for Males and Females Combined

($N = 300$)

Variables	1	2	3	4	5	6	7	8	9	10	*M*	*SD*
OARS:												
1. Materialistic											4.80	2.48
2. Competitive	27										6.65	2.01
3. Technical	03	10									5.03	2.28
4. Humanitarian	−05	21	12								5.37	2.58
WSQ:												
5. Managerial-Aggressive	−10	38	−01	27							32.16	4.83
6. Status-Need	26	14	−14	−21	−06						15.97	4.80
7. Organization-Need	16	16	31	22	29	−10					28.70	4.85
8. Working Conditions	12	03	−08	−22	00	28	02				14.77	3.70
JIR:												
9. Achievement-Need	−43	−07	12	11	16	−14	−09	−16			9.85	3.16
10. Intrinsic Job Interest	−35	−07	03	23	12	−22	06	−18	39		7.47	3.22
11. Job Autonomy	18	−14	−02	02	−14	−13	18	11	−27	−09	10.47	3.22

Note.—Decimal points omitted for correlation coefficients. At .05 level of significance $r = .15$; at .01 level of significance $r = .23$.

TABLE 2

FACTOR LOADINGS OF OCCUPATIONAL ATTITUDE RATING SCALES (OARS), WORK SATISFACTION QUESTIONNAIRE (WSQ), AND JOB INCENTIVE RANKINGS (JIR) VARIABLES

(N = 300)

Variables	Rotated Factors					h^2
	A	B	C	D	E	
OARS:						
1. Materialistic	56	00	03	05	−05	51
2. Competitive	03	−08	52	04	−07	54
3. Technical	−02	−03	00	70	01	54
4. Humanitarian	−04	−40	40	−01	−03	35
WSQ:						
5. Managerial-Aggressive	−39	05	74	−07	34	59
6. Status-Need	04	39	−04	−03	−05	36
7. Organization-Need	03	04	44	35	48	51
8. Working Conditions	−03	50	02	02	30	30
JIR:						
9. Achievement-Need	−62	00	06	15	−07	51
10. Intrinsic Job Interest	−39	−16	15	02	05	32
11. Job Autonomy	37	−05	−04	−05	31	36

Note.—Decimal points omitted.

scales and the scatterplots of five randomly selected pairs of scales, the assumptions of normality and linearity of regression seem tenable.

Table 2 lists the rotated factor loadings and communalities for the 11 variables. Factors A and B have both positive and negative loadings of sufficient magnitude to make them bipolar. The highest absolute loadings are on Factors C and D and the lowest on Factors B and E, with Factor A intermediate. Six of the scales (Materialistic, Competitive, Technical, Status-Need, Achievement-Need, and Intrinsic Job Interest) load on only one factor, whereas the other five (Humanitarian, Managerial-Aggressive, Organization-Need, Working Conditions, and Job Autonomy) relate to two or more factors. With the exception of the negative pole of Factor B, as defined by the Humanitarian scale, the factors are composites, being formed by combinations of two or more scales.

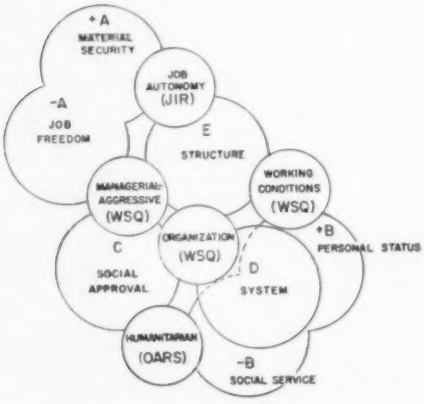

FIG. 1. Vocational motivation factors.

Figure 1 presents graphically the various factors and some of the scales which load on two or more of them. The scales which correlate with only one factor are not shown, in order to simplify the diagram, but they would be within the larger circles. The circles depict the total variance in a factor or scale, and the amount of overlap between circles represents the approximate percentage of common variance. The connected circles for Factors A and B indicate that they are bipolar. Factor D is superimposed upon Factor B, but the two factors are actually independent.

Discussion

To facilitate the interpretation of the results and to make explicit the bases for inferences about the meaning of the factors refer to Figure 1 and Table 3. In the latter the columns (factors) and rows (scales) correspond to those in Table 2, but the entries are the items from the various vocational motivation scales. For a given scale, items were classified according to the factors which they seemed to define. Thus, reading down a column provides a summary of the item content for a factor and suggests some hypotheses about the variable it measures.

The factors were named, as much as possible, to indicate the dimension of vocational motivation they define, instead of the type of work which satisfies a particular need or value. Three general dimensions, each with

TABLE 3

OARS, WSQ, and JIR Scale Items Classified by Vocational Motivation Factors

Scales	Factors				
	A. Material Security vs. Job Freedom	B. Personal Status vs. Social Service	C. Social Approval	D. System	E. Structure
1. Materialistic	(Positive) Economic security Earn big money Regular salary Earn fat bonuses Lots of overtime Quick financial return				
2. Competitive			Be the "Top Man" Room for self-expression Be able to assert self Looked up to in the community Heavy responsibilities Provide leadership Make a name for self Recognition		
3. Technical				Using math Attention to details Working with figures A methodical approach Working with theory Using hands and tools skillfully Working in privacy	
4. Humanitarian		(Negative) Being patient with people Hearing people's troubles Be of service to others Change people for better Help others help themselves	Initiate group action Influence people's lives Improve social conditions Right social wrongs Help the unfortunate		

TABLE 3—(*Continued*)

Scales	Factors				
	A. Material Security vs. Job Freedom	B. Personal Status vs. Social Service	C. Social Approval	D. System	E. Structure
5. Managerial-Aggressive	(Negative) A job where I do *not* work under instructions Work where the duties change frequently		Directing, controlling, and planning activities of others Work where I influence opinions and attitudes of others A job where I work under stress A job where I express my personal ideas and feelings		Work which keeps me very busy all of the time
6. Status-Need		(Positive) Recognition from others Travel frequently High salary Live in large city			
7. Organization-Need			Working closely along with others Working with people who have similar interests	Make judgments according to measurable standards A job where I must attain exact standards	Work which keeps me very busy all of the time A job where the work must be performed according to a set time schedule
8. Working Conditions		(Positive) A job where I work indoors all of the time A job with no physical activity Live in large city			Work where the duties do *not* change frequently
9. Achievement-Need	(Negative) Freedom to assume responsibility			Opportunity to learn new skills	
10. Intrinsic Job Interest		(Negative) Friendly fellow workers			
11. Job Autonomy	(Positive) Good job security				Good job security

Note.—The Managerial-Aggressive and Organization-Need scales of the Work Satisfaction Questionnaire have the item "Work which keeps me busy all of the time" in common.

specific aspects, emerged. As discussed below, these were security, status, and service:

Factor A—Material Security vs. Job Freedom. The loadings of the so-called Job Autonomy (JIR) scale and the Materialistic (OARS) scale define the positive pole of this factor as a motivation for the extrinsic rewards rather than the intrinsic satisfactions of work. The emphasis in the negative pole is upon those aspects of the Achievement-Need (JIR) and Managerial-Aggressive (WSQ) scales which pertain to freedom in work, particularly as expressed in varied job duties and minimal supervision. The two poles of

this factor closely correspond to the extrinsic and intrinsic sources of job satisfaction suggested by Ginzberg et al. (1951).

Factor B—Personal Status vs. Social Service. The positive pole of this factor, as defined by the Status-Need (WSQ) and Working Conditions (WSQ) scales, reflects an egocentric need for the status which comes from a highly remunerative, prestigeful, white collar job. Unlike the status acquired through the admiration and acceptance of others, which Factor C (Social Approval) assesses, the essence of the need for personal status is affluence rather than influence. The negative pole is the familiar "helping others" motivation, but always on an intimate, personal level, not through organized activities and large-scale programs. Evidently, there are two distinct modes for the expression of humanitarian motives: one is in face-to-face individual relationships, and the other is through group enterprises, as in Factor C (Social Approval).

Factor C—Social Approval. Unidentified in previous work, but intriguing as a dimension of vocational motivation, this factor appears to represent a fusion of attitudes inherent in the Protestant Ethic and American Capitalism (Whyte, 1956). It reflects the appeal of those social rewards, e.g., recognition as a leader and community servant, which stem from the initiation and implementation of political and service programs through aggressive executive action. The prototype of the person motivated by "social approval," perhaps because of guilt feelings about his material well-being, is the successful businessman who organizes charities and establishes funds in the public interest.

Factor D—System. Best defined by the Technical (OARS) scale, which encompasses the aspects of precision and skill contributed by the Organization-Need (WSQ) and Achievement-Need (JIR) scales, this factor was named "System" to identify its main characteristics: order, planfulness, detail, perfectionism, and impersonality. These characteristics define an orientation which contrasts sharply with that reflected in Factor E (Structure). The difference is much like that which exists between the "methodical" man and the "mechanical" man: one creates plans, and the other carries them out.

Factor E—Structure. Other designations, such as "Organizational Security," which emphasize the importance of duties and tasks, routines, and schedules as means of anxiety reduction (Kates, 1950), also capture the meaning of the scales that form this factor. In this instance, as contrasted with Factor A (Material Security), the formality of an impersonal organizational structure, rather than the accumulation of worldly goods, serves to provide a source of security.

These interpretations of the factors are necessarily only suggestive and tentative, but they seem meaningful and provide hypotheses for further research.

With respect to the OARS, WSQ, and JIR scales, their intercorrelations and factorial loadings provide some insight into both their composition and potential usefulness:

1. Occupational Attitude Rating Scales. The Materialistic and Technical scales are factorially pure and measure what their names imply, although they might better be identified by more psychological and motivational terms, such as the needs for "Material Security" and "System." The Competitive scale is also factorially simple, but it seems to measure less a need for "Personal Status," as Hammond (1956) suggests, than a need for "social status," as indicated by its loading on Factor C (Social Approval). Unlike the others, the Humanitarian scale is factorially complex: one part pertains to assisting people through empathic understanding, the other represents helping them through group action.

2. Work Satisfaction Questionnaire. With the exception of the Status-Need scale, which assesses a desire for prestige and self-aggrandizement, the scales in this inventory are nonspecific and nonunitary. The Managerial-Aggressive scale breaks down into three distinct parts—job freedom and variety, controlling and directing, and rapid work pace—which support the appropriateness of its name. Similarly, the Organization-Need scale subdivides into various components, but only the items which load on Factor E (Structure) are consistent with the scale's title. The Working Conditions scale, which Astin

(1958) refrained from naming, assesses essentially what its title connotes and has sufficient promise to recommend its use in further research.

3. Job Incentive Rankings. The incentive rankings have two major shortcomings. First, the names of the scales are largely inappropriate to express what they measure. The Achievement-Need scale pertains more to job autonomy than to "getting ahead." Likewise, the Intrinsic Job Interest scale concerns concomitant work satisfactions rather than satisfactions derived from the work per se. Also, the Job Autonomy scale measures security needs, not the desire for freedom in work. And, second, the scales are too restricted in their coverage of the vocational motivation domain. They are not comprehensive enough to represent more than one or possibly two of the seven factors derived from the analysis. Probably the best set of scales to define all of the factors, for both males and females, since there are no sex differences, is the following: Factor A—Materialistic (OARS) and Achievement-Need (JIR), Factor B—Status-Need (WSQ) and Humaniatrian (OARS), Factor C—Managerial-Aggressive-(WSQ), Factor D—Technical (OARS), and Factor E—Organization-Need (WSQ).

As far as theories of vocational motivation are concerned, they propose various drives, instincts, and needs as basic in explaining why individuals make differential choices of occupations and seek dissimilar satisfactions in work. Among the energizing states and conditions which supposedly elicit and produce individual differences in vocational behavior are the aggressive and destructive drives (Menninger, 1942), the mastery instinct (Hendrick, 1943a, 1943b), the work and pleasure orientations (Ginzberg et al., 1951), and the physiological and self-actualization needs (Roe, 1956). To assess the value of these constructs as definitions of independent dimensions of vocational motivation it is necessary to translate them into operational terms and to determine their interrelationships. The OARS, WSQ, and JIR scales measure a number of the variables specified in the theories, and their intercorrelations and factorial loadings reveal the extent to which the variables covary. The findings agree with the use of several theoretical constructs to account for individual differences in vocational motivation: there is no one general vocational motivation factor. More specifically, they suggest that the needs for security, status, and service are unique vocational motives and that Ginzberg's theory of extrinsic, concomitant, and intrinsic work satisfactions is the most meaningful one as a frame of reference for these concepts.

In further research on the factors as definitions of vocational motivation, one of the most important problems concerns the demonstration of their construct validity. The present study establishes the factors as unique variables, and suggests hypotheses about their meaning, but it offers no independent evidence of what they measure. Two kinds of additional studies are needed: one on the relationship of the factors to nontest indices of motivation, and the other on the possible drive properties of the factors, such as their efficacy in the facilitation of a wide variety of responses and in the learning of new responses (Brown, 1961). Once there is support for the factors as measures of motivation, rather than habits or response tendencies, research on their relationships to vocational interests (Darley & Hagenah, 1955), occupational membership (Roe, 1956), and job satisfaction (Ginzberg et al., 1951) will be possible.

Summary

Recent theoretical emphases upon the relationship of motivational variables to occupational choice and vocational adjustment have resulted in the development of several instruments designed to measure such vocational motives as competition, achievement, and independence. Since the various scales of these newly constructed questionnaires and inventories either differentiate similar educational and vocational groups or have comparable item content, relationships among them are suggested. The present study was designed to test for the existence of these interscale relationships and to formulate factor analytic definitions of vocational motives.

From an 11-variable intercorrelation matrix based upon test data for 300 subjects (150 males and 150 females), five factors were extracted, following the complete centroid method. After rotation, the factors

were identified as follows: A—Material Security vs. Job Freedom, B—Personal Status vs. Social Service, C—Social Approval, D—System, and E—Structure. These factors appeared to provide a comprehensive survey of the dimensions of vocational motivation usually mentioned in theories of occupational choice and vocational adjustment. Intrepretations of the factors were made, and the implications of the findings for theory construction were discussed. Some possible problems for future research on the construct validity of the factors and their relationships to other vocational variables were outlined.

REFERENCES

ASTIN, A. W. Dimensions of work satisfaction in the occupational choices of college freshmen. *J. appl. Psychol.*, 1958, **42**, 187–190.

BENDIG, A. W., & STILLMAN, EUGENIA L. Dimensions of job incentives among college students. *J. appl. Psychol.*, 1958, **42**, 367–371.

BROWN, J. S. *The motivation of behavior*. New York: McGraw-Hill, 1961.

DARLEY, J. G., & HAGENAH, THEDA. *Vocational interest measurement*. Minneapolis: Univer. Minnesota Press, 1955.

GINZBERG, E., GINSBURG, S. W., AXELRAD, S., & HERMA, J. L. *Occupational choice*. New York: Columbia Univer. Press, 1951.

HAMMOND, MARJORIE. Occupational attitude rating scales. *Personnel guid. J.*, 1954, **32**, 470–474.

HAMMOND, MARJORIE. Motives related to vocational choices of college freshmen. *J. counsel. Psychol.*, 1956, **3**, 257–261.

HAMMOND, MARJORIE. Attitudinal changes of "successful" students in a college of engineering. *J. counsel. Psychol.*, 1959, **6**, 69–71.

HENDRICK, I. The discussion of the "instinct to master." *Psychoanal. Quart.*, 1943, **12**, 561–565. (a)

HENDRICK, I. Work and the pleasure principle. *Psychoanal. Quart.*, 1943, **12**, 311–329. (b)

KATES, S. L. Rorschach responses related to vocational interests and job satisfaction. *Psychol. Monogr.*, 1950, **64**(3, Whole No. 309).

LAWLEY, D. N. The maximum likelihood method of estimating factor loadings. In G. Thomson (Ed.), *The factorial analysis of human ability*. (5th ed.) New York: Houghton Mifflin, 1951.

MENNINGER, K. A. Work as sublimation. *Bull. Menninger Clin.*, 1942, **6**, 170–182.

ROE, ANNE. *The psychology of occupations*. New York: Wiley, 1956.

WHYTE, W. H. *The organization man*. New York: Simon & Schuster, 1956.

(Received November 25, 1960)

Journal of Applied Psychology
1961, Vol. 45, No. 5, 338–344

INDIVIDUAL AND GROUP CORRELATES OF ATTITUDES TOWARD WORK-RELATED CHANGE[1]

DON A. TRUMBO

Kansas State University

This report is one in a series describing a set of studies on technological change. An earlier report (Jacobson, Trumbo, Cheek, & Nangle, 1959) described the research setting, sample, and data gathering procedures and presented selected findings regarding employee perceptions of and attitudes toward a specific technological change event.

In the present study, individual and group correlates of attitudes toward change were explored for the purpose of refining a number of hypotheses regarding those factors which condition employees attitudes toward work-related change. The term "work-related change" serves to distinguish the more general attitude object of the present study from the specific change event, reactions to which were described in the initial report.

Prior research on response to industrial change has focused on the work group as the unit of study. Resistance to change has been defined in terms of group performance measures and turnover rates (Bavelas, 1947; Coch & French, 1948). The effectiveness of "group participation" methods of administering change, as indicated in the results of these studies, strongly suggests that the leadership attitudes and behavior of the supervisor, in his role as change agent, may condition group attitudes toward change.

While these studies provide insight into the factors which facilitate the acceptance of specific change events, no research known to this writer has systematically investigated the attitudes of employees toward change as a general work-related phenomenon. Neither correlates of individual change attitudes nor situational factors asociated with differences among attitudes in work groups have been identified.

In the present study it was assumed that attitudes toward change could be meaningfully related to personal data items and other indices of employee needs and abilities, and that from these relationships more specific hypotheses about the underlying need structure and conditioners of change attitude could be derived. It was also assumed that the social psychological climate of the work group, as reflected in measures of supervisor's attitudes and group cohesiveness, conditions the change attitudes of the group.

At the level of the individual employee, the following general hypotheses were stated:

Attitudes toward change are positively related to the job investment of the employee. Job investment was defined in terms of the employee's identification with the work force, the economic need for the job, and length of service with the company. In addition both age and sex of the employee were felt to be indicative of differences in job investment, with the older employees having a greater investment than younger employees and men having a greater investment than women. Therefore, the personal data items of age, sex, length of service, and a job involvement (JI) index were considered together as indices of job investment. The prediction of a positive relationship between these items and attitudes toward change was based on the assumption that job changes necessitate readjustments which would have more negative valence for the employees with little job investment than for those with a greater investment in their jobs.

Secondly, it was hypothesized that attitudes toward change are positively related to the capacity for adjustment to change. Capacity for adjustment was defined in terms

[1] This paper is based on portions of the writer's doctoral dissertation (Trumbo, 1958). The writer wishes to acknowledge the guidance and assistance of James S. Karslake, major professor, and of Eugene Jacobson, Milton Rokeach, and Donald M. Johnson. This study was conducted as one of a series of projects on technological change supported by the Labor and Industrial Relations Center, Michigan State University, under the research directorship of Jack Stieber. Jacobson served as project director.

of general ability, amount of formal education, and freedom from job anxiety.

At the work group level, relationships were predicted between three measures of supervisory attitudes and group attitudes toward change. In view of the supervisor's role as administrator of change and in light of the evidence for the effectiveness of group participation methods of administering change, positive relationships were predicted (*a*) between the change attitudes of the supervisor and group attitudes toward change, and (*b*) between the human relations attitudes of the supervisor and the change attitudes of the group. Prior research also provided a rationale for predicting a negative relationship between the authoritarianism of the supervisor and group change scores.

Finally, reasoning that change may present a threat to the satisfaction and security obtained in cohesive work groups, it was predicted that cohesiveness would be negatively related to attitudes toward change in the work group.

Since there is little in the way of formal theory regarding the phenomenology of work-related change, these predictions were tentative and the study largely an exploratory one.

Procedures

Data were obtained from questionnaires and from the personnel files on 46 supervisory and 232 nonsupervisory personnel, approximately 85% of the home office staff of an insurance company engaged in "office automation" changes. Of the nonsupervisory sample, 79% were women; the median age was 30 years (interquartile range: 23 to 42 years), and the median length of service was 15 months. Eighty-two percent were high school graduates, and an additional 10% were college graduates. A more detailed description of the sample, the questionnaires, and data collection procedures is available in the previous report (Jacobson et al., 1959) and need not be repeated here.

The Change Scale. The major index of employee attitudes toward work-related change consisted of the following nine Likert-type items from the questionnaire:

1. The job that you would consider ideal for you would be one where the way you do your work: (five choices from "is always the same" to "change a great deal").
2. If I could do as I pleased, I would change the kind of work I do every few months.
3. One can never feel at ease on a job where the ways of doing things are always being changed.
4. The trouble with most jobs is that you just get used to doing things in one way and then they want you to do them differently.
5. I would prefer to stay with a job that I know I can handle than to change to one where most things would be new to me.
6. The trouble with many people is that when they find a job they can do well, they don't stick with it.
7. I like a job where I know that I will be doing my work about the same way from one week to the next.
8. When I get used to doing things in one way it is disturbing to have to change to a new method.
9. It would take a sizeable raise in pay to get me to voluntarily transfer to another job.

Each of the remaining eight items had five alternatives from "I strongly agree," to "I strongly disagree."

Responses to these items were coded serially from 1 to 5 with high item scores indicating favorable change attitudes. Summated scores ranged from 11 to 45 with a mean of 28.03 ($\sigma = 6.64$). Comparigroups ($N_u = N_l = 61$) from the composite score distribution yielded t values from 3.5 to 13.5 with a median $t = 12.1$ ($p < .01$ for each item). All items were monotonically related to total scores. The corrected odd-even reliability coefficient was .79. Hence, the items met the criteria for a Likert scale and were labeled the Change Scale.

While the Change Scale had face validity as a measure of generalized attitudes toward change, evidence of logical validity was sought in a comparison of scale scores with responses to questions about specific past, current, and anticipated future change events. If the scale were a valid measure of attitudes toward change, high scores would be associated with favorable responses when much change was perceived and unfavorable responses when little or no change was perceived. Conversely, low Change Scale scores would be predictive of unfavorable responses when much change was perceived and favorable responses when little or no change was perceived.

Four pairs of Likert-type items in the questionnaires permitted a test of these predictions. The first item in each pair indicated the amount of change perceived ("Have machines changed the nature of your job in the past two years?" "What effect did the change-over [to the new computer] have on your job?" "In general, how much change takes place from time to time in the way you do your present job?" and "Do you think the computer will affect your job in the next year or two?"). Each of these items was followed by the same five-choice item: "How do (did) you feel about this?" Alternatives to this item ranged from "like it very much" to "dislike it very much."

Employees were dichotomized into "high change" and "low change" groups on the basis of responses

to the first item in each pair. These groups were then subdivided into "like" (L), "indifferent" (I), and "dislike" (D) categories by their responses to the second item in each pair. Differences among the mean Change Scale scores for the L, I, and D subgroups of the "high change" employees were significant and in the predicted direction for each of the four items. For "low change" employees, the differences were consistently in the predicted direction, but significant for one item only.

These comparisons indicate that Change Scale scores were predictive of attitudes toward specific change situations, particularly when the employee perceived or anticipated relatively extensive changes in his own job.

Job investment. Data on length of service, age, and sex were obtained from the questionnaires and from the personnel files. The JI index was a composite score based on two questionnaire items dealing with future identification with the work force ("What do you *hope* to be doing five years from now?" and "What do you *really expect* to be doing five years from now?") and two items indicating the economic significance of the job ("Are you the main wage earner in your household?" and "Could your household live adequately if you were not working?"). Item responses were dichotomized and scored "0" (low involvement) or "1" (high involvement). Intercorrelations among the four items ranged from .38 to .89.

Capacity for job adjustment. Indices of the employee's capacity for adjustment to job changes included Personnel Test scores (Wonderlic & Hovland, 1939), years of formal education, and a single item job anxiety index. Personnel Test scores were available from the personnel files for 181 of the 232 employees. Job anxiety was indicated by responses to the item: "Do things here at work ever make you feel 'jumpy' or nervous?" (six alternatives from "never" to "very often").

Supervisory attitudes. The Change Scale was included in the supervisors' form of the questionnaire, together with the Leadership Inventory (Nelson, 1949) and an abbreviated (11-item) form of the Dogmatism Scale (Rokeach, 1956). Scores on these three scales served as measures of supervisory attitudes toward change, toward human relations practices, and toward authority, respectively.

The Leadership Inventory consists of 50 two-choice items with alternatives designed to reflect four idealized patterns of leadership behavior (Autocratic, Bureaucratic, Idiocratic, and Democratic). Item stems are presented twice, once with A and D alternatives, and again with B and I alternatives. In the present study, items were combined into 25 four-choice items. A Human Relations score was obtained by summing the number of Idiocratic and Democratic alternatives chosen.

The 11 items for the abbreviated Dogmatism (D) Scale were selected on the basis of item analysis data presented elsewhere (Rokeach, 1956). The D scale was developed by Rokeach (1956) as a measure of general authoritarianism. Conceptually, dogmatism is, in part, an attitude toward authority. High dogmatic individuals are conceived of as being more accepting or rejecting of others depending upon whether they agree or disagree with their own beliefs. The dogmatic personality is at the same time aggressive toward peers and subordinates and submissive toward persons in positions of authority over him (Trumbo, Rokeach, & Gladin, 1957). Such attitudes toward authority and interpersonal relations may be important conditioners for the behavior and attitudes of the subordinates.

Group cohesiveness. The five-item index of group cohesiveness used by Seashore (1954) was included in the questionnaire. These items were designed to measure the extent to which an individual perceives himself as belonging to a group, desires to remain in the group, and perceives his group to be superior to others.

The predicted relationships were tested by product-moment correlations or by tests of differences between means, depending on the nature of the available data.

Results

Attitudes toward Change and Indices of Job Investment

Data on age, sex, length of service, and scores on the JI index were obtained for nearly all of the 232 nonsupervisory employees. Slight variations in N resulted when questionnaire items were omitted or personnel records were incomplete. The results are summarized in Table 1.

Sex. Change Scale scores were available for 47 male and 178 female employees. The mean score for men was 31.4 and for women 27.1. The difference of 4.3 points yielded a t value of 3.98 which was significant ($p < .001$).

Age. Information on date of birth was recorded for 217 of the 232 employees. On the basis of the age distribution, employees were grouped into six age level categories. A mean Change Scale score was computed for each age group. An F test of the differences among these means was not significant ($F = 1.41$ with $5/211$ df; $p > .05$).

Length of service. As indicated in Table 1, data were available on length of service with the company for 209 employees. Because of a rather high turnover rate, particularly among the younger female employees, the distribution of length of service was highly skewed. Consequently, the distribution was dichotomized at the median of 15

TABLE 1

SUMMARY OF CORRELATES OF INDIVIDUAL CHANGE SCALE SCORES

Variable	Group	*N*	Mean Change Scale Score	Test
Sex	M	47	31.4	$t = 3.98^{**}$
	F	185	27.1	
Age (in years)	17–20	31	26.3	
	21–25	58	29.0	
	26–30	28	30.6	
	31–35	36	28.1	
	36–45	34	27.6	
	45+	30	26.1	$F = 1.41$
Length of Service	+15 months	112	28.2	$t = .26$
	−15 months	97	28.0	
Job Involvement		229		$r = .13$
Personnel Test		181		$r = .28^{**}$
Education:				
Less than High School		17	26.8	
High School		126	27.0	
Less than College		57	27.7	
College		23	33.8	$F = 6.94^{**}$
Job Anxiety		229		$r = -.17^{**}$

** $p < .01$.

months into long- and short-term groups. Comparison of the mean Change Scale scores for these two groups failed to indicate any difference in attitudes toward change ($t = .26$).

Job involvement. The JI index yielded five score intervals from 0 to 4. Scores for 229 employees were essentially rectangularly distributed over the five score intervals, therefore, to test for relationship, a mean Change Scale score was computed for employees at each JI score interval. An *F* test of the variance among these means was not significant ($F = 1.36$ with $4/224$ *df*; $p > .05$). The direction of the relationship between JI and Change Scale scores was positive, however, as indicated by a product-moment correlation of .13 (*ns*).

Attitudes toward Change and Indices of the Capacity of Job Adjustment

Each of the indices of the employee's capacity for adjustment to changes in the job was found to be predictive of change attitudes. The results are presented in the lower portion of Table 1.

General ability. A product-moment correlation of .28 between scores on the Wonderlic Personnel Test and the Change Scale was significantly different from zero (with 179 *df*, $p > .001$).

Amount of education. Data on years of formal education were obtained for 223 employees, who were subseqently grouped at four levels of education: less than high school ($N = 17$), high school ($N = 126$), some college ($N = 57$) and college and graduate work ($N = 23$). An *F* test of the differences among the mean Change Scale scores of these four groups yielded a significant *F* of 6.94 (with $3/219$ *df*, $p < .01$). Differences were consistent in direction, indicating a positive, monotonic relation between amount of education and attitudes toward change. The mean for the college and graduate group (33.8) was significantly different from each of the other means, but none of the remaining differences was significant.

Job anxiety. Responses to the single item index of job anxiety ("Do things here at work ever make you feel 'jumpy' or nervous?") were distributed as follows: never ($N = 18$), very seldom ($N = 49$), seldom ($N = 54$), sometimes ($N = 91$), quite often ($N = 14$), very often ($N = 3$). Scores were assigned serially from 1 to 6 to these response categories. A product-moment correlation between these scores and Change Scale scores yielded a coefficient of $-.17$ which was significantly different from zero (with 227 *df*, $p < .01$).

Except for differences between men and women, the four variables (sex, age, length of service, and job involvement) assumed to be indicative of differences in the job investment of the employee failed to be related significantly to attitudes toward change. Consequently, little support is provided for the prediction that the employee with greater job investment would have more favorable attitudes toward change. On the other hand, less favorable attitudes among the female employees may indicate that change is perceived as threatening to social aspects of the job, which women rate as relatively more important than do men (Herzberg, Mausner, Peterson, & Capwell, 1957). Support for this interpretation was found in the analysis of situational factors presented below.

Indices of general ability, amount of education,

and freedom from job worries were each found to be predictive of more favorable attitudes toward change. These variables appear to reflect three aspects of the employee's capacity to adjust to changes in his work. The results consistently support the prediction that attitudes toward change are related to the capacity to adjust to change situations.

Attitudes toward Change in the Work Group

Of the 232 nonsupervisory employees, 167 were identified with 21 work groups from each of which three or more employees had completed the questionnaires. The work groups, defined in terms of common first level supervision, ranged in size from 3 to 27 with a median of four members represented in the questionnaire data.

Basic to the decision to study change attitudes at the group level was the assumption that such attitudes are associated with group membership. To test this assumption, a simple analysis of variance was performed on the Change Scale scores of the 21 work groups. The resulting F value of 2.64 was significant (with 20/146 df, $p < .01$), indicating that employees within work groups are relatively homogeneous with respect to attitudes toward change. A Bartlett's test of homogeneity of variance was not significant ($\chi^2 = 3.77$), supporting the conclusion that the significant between-group variance was indicative of differences among the group means.

These results demonstrate both that change attitudes are associated with group membership and that groups can be meaningfully described in terms of attitudes toward change. Consequently, Group Change scores were computed for the 21 groups as the mean of the Change Scale scores of group members.

Group cohesiveness and group change scores. A Cohesiveness score was obtained for each of the 21 groups by averaging the scores of group members on the five-item index. An F test of the between group variance yielded an F of 3.91 (with 20/146 df, $p < .01$), confirming Seashore's (1954) finding that cohesiveness is a meaningful dimension when applied to industrial work groups, and supporting the decision to use the mean scores as measures of group cohesiveness.

The predicted negative relationship between Cohesiveness and Group Change scores was tested by computing a product-moment correlation between the two sets of scores. The results, in Table 2, were consistent with the prediction. The coefficient of $-.50$ was significantly different from zero (with 20 df, $p < .01$).

Supervisors' attitudes and group attitudes toward change. Results of the correlations between three indices of supervisors' attitudes and Group Change scores are presented in Table 2.

The Human Relations (HR) scores from the Leadership Inventory provided a logically valid index of employee-oriented supervisory attitudes. Since, as indicated earlier, a social climate conducive to group participation and decision making appears to reduce overt resistance to change, it was predicted that HR scores would be positively related to Group Change scores. The results failed to support this prediction, however. The coefficient of .04 obtained between these two sets of scores was obviously not significant.

The supervisors' Change Scale scores served as the second measure of supervisory attitudes. Assuming the supervisor's role as an administrator of change to be an important conditioner of group change attitudes, a positive relationship was expected between Change Scale scores of the supervisor and those of the group. The coefficient of .41 between these two sets of scores supports this prediction (with 20 df, $p < .05$).

TABLE 2

SUMMARY OF CORRELATES OF GROUP CHANGE SCORES AND THEIR INTERCORRELATIONS

(N = 21 groups)

	Change Scale (supervisors)	Cohesiveness	Human Relations score (supervisors)	Dogmatism (supervisors)
Change Scale (group)	.41*	−.50**	.04	−.38*
Change Scale (supervisors)		−.65**	−.12	.01
Cohesiveness			−.18	.20
Human relations score (supervisors)				.12

* $p < .05$ (two-tailed test).
** $p < .01$ (two-tailed tests).

The final measure of supervisors' attitudes was the abbreviated Dogmatism (D) Scale. As indicated earlier, a negative relationship was predicted between supervisors' Dogmatism scores and Group Change scores. The results are consistent with this prediction. The correlation between these two sets of scores was $-.38$, which is significantly different from zero ($p < .05$ with 20 df).

Discussion

Failure to find a significant relationship between the human relations attitudes of the supervisors and group attitudes toward change was rather unexpected in view of the popular belief about the effectiveness of "employee oriented" supervision. One explanation of this lack of relationship may be found in the restricted range of human relations attitudes among the supervisors. In this connection, it was found that the standard deviation of HR scores of the 21 supervisors in the group analysis was considerably, though not significantly, less than that of the population of 46 supervisors from which they were drawn ($\sigma_{(21)} = 6.79$, whereas $\sigma_{(46)} = 11.62$).

Secondly, it is possible that, while human relations attitudes of the supervisor condition certain attitudes of their subordinates, it should not be assumed that they condition all work-related attitudes. However, no support for this explanation was found. HR scores were not significantly related to either group cohesiveness scores or summated scores on a 15-item job satisfaction check list.

The negative relationship between Cohesiveness and Group Change scores may indicate that change is perceived as a source of threat to informal group structure and to the satisfaction of social needs provided for in a cohesive work group. This interpretation, which was given earlier to account for the less favorable attitudes of female workers, appears to be supported by both findings.

A more adequate understanding of the phenomenology of industrial change requires an analysis of employee needs in relationship to change attitudes and of the relationship of this attitude dimension to other job-related attitudes. An employee's response to change is probably conditioned by his perception of the way in which the effects of change relate to his needs. If change as a general phenomenon is to be accepted, its effects must be perceived as generally more rewarding than unrewarding, that is, they must provide need satisfaction.

What are the effects of change which may condition change attitudes? The most obvious effect is that change provides variety in the job routine. Thus, favorable attitudes toward change may reflect a need for variety or avoidance of repetition. Change may also enhance the opportunities for promotion thereby providing satisfaction for status needs. Finally, skill requirements and responsibility demands may be increased as a result of change, thereby providing greater challenge and perhaps greater self-expression in one's work.

These speculations about the need structure underlying attitudes toward change gain some support from the following evidence: Positive relationships were found between Change Scale scores and the extent to which employees liked reported *increases* in variety, skill requirements, responsibility demands, and chances for promotion in their work. These results suggest that employees may favor change because they perceive it as a means to greater variety, status, and self-expression at work.

An alternative interpretation would be that readiness for change is simply a manifestation of job dissatisfaction. That is, an employee may welcome change because he is dissatisfied with his job in general or with specific aspects of the job, such as fellow employees or supervision. Evidence bearing on this interpretation was negative, however. The correlation between Change Scale scores and summated scores on the 15-item Job Satisfaction (JS) check list was .07. Furthermore, upper and lower 27% groups from the Change Scale distribution did not differ significantly on any of the 15 items of the JS check list.

In review, the following tentative conclusions may be presented as hypotheses for further research. Employees may perceive change as a source of threat to social aspects of their job, as indicated both by the less favorable attitudes of females and by the negative correlation between group cohesiveness and Group Change scores. Situational

factors, manifested in supervisory attitudes and group cohesiveness, appear to be important conditioners of change attitudes in the work group. Individual differences in attitudes toward change may reflect differences in the capacity to adjust to change situations. Finally, readiness for change may be related to a complex of needs including variety, status, and self-expression at work.

Summary

This report presented some of the findings of a study of the correlates of employee attitudes toward change as a general job-related phenomenon. Questionnaire and personnel file data were obtained for 232 nonsupervisory and 46 supervisory personnel of a medium sized midwestern insurance company involved in office automation. Attitudes toward change were measured with a nine-item Change Scale included in the questionnaires. Evidence for the reliability and logical validity of the Scale were presented.

The results may be summarized as follows:

1. Female employees scored significantly lower on the Change Scale than males. However, since none of the other indices of job investment (age, length of service, and Job Involvement index) was predictive of change attitudes, the sex difference was interpreted in terms of the differential importance assigned to various aspects of the job by men and women. Since women consistently rate social aspects of the job as relatively more important than do men, it was hypothesized that unfavorable attitudes toward change may reflect a perceived threat to informal social structure.

2. Change Scale scores were positively related to Wonderlic Personnel Test scores, amount of education, and freedom from job anxiety, supporting the view that change attitudes are related to the capacity to adjust to changes.

3. Attitudes toward change were found to be associated with work group membership. Change Scale scores within groups being relatively more homogeneous than among groups.

4. Group cohesiveness was negatively related to Group Change scores. This finding gave support to the view that less favorable attitudes toward change may indicate that change poses a threat to the satisfaction of social needs through informal social structure.

5. Supervisors' attitudes toward change were positively related to Group Change scores, while supervisors' scores on a measure of authoritarianism were negatively related to Group Change scores. However, supervisors' scores on an index of human relations attitudes were unrelated to attitudes of the group toward change.

6. Among employees who perceived increases during the preceding year in variety, skill and responsibility demands, and chances for promotion, approval of these increases was associated with higher Change Scale scores than indifference or disapproval. This evidence provided tentative support for the view that readiness for change is related to employee needs for variety, status, and self-expression at work.

REFERENCES

Bavelas, A. Cited by K. Lewin. Group decision and social change. In T. M. Newcomb & E. L. Hartley (Eds.), *Readings in social psychology*. New York: Holt, 1947.

Coch, L., & French, J. R. P., Jr. Overcoming resistance to change. *Hum. Relat.*, 1948, **1**, 512–532.

Herzberg, F., Mausner, B., Peterson, R., & Capwell, D. *Job attitudes: Review of research and opinion*. Pittsburgh, Pa.: Psychological Services of Pittsburgh, 1957.

Jacobson, E., Trumbo, D., Cheek, G., & Nangle, J. Employee attitudes toward technological change in a medium sized insurance company. *J. appl. Psychol.*, 1959, **43**, 349–354.

Nelson, C. W. The development and evaluation of a leadership attitude scale for foremen. Unpublished doctoral dissertation, University of Chicago, 1949.

Rokeach, M. Political and religious dogmatism: An alternative to the authoritarian personality. *Psychol. Monogr.*, 1956, **70**(18, Whole No. 425).

Seashore, S. E. *Group cohesiveness in the industrial work group*. Ann Arbor: University of Michigan, Institute for Social Research, 1954.

Trumbo, D. A. An analysis of attitudes toward change among the employees of an insurance company. Unpublished doctoral dissertation, Michigan State University, 1958.

Trumbo, D., Rokeach, M., & Gladin, L. A validation study with high and low dogmatic groups. Paper read at Midwestern Psychological Association, Chicago, May 1957.

Wonderlic, E. F., & Hovland, C. I. The Personnel Test: A restandardized abridgment of the Otis S-A test for business and industrial use. *J. appl. Psychol.*, 1939, **23**, 685–702.

(Received November 25, 1960)

Journal of Applied Psychology
1961, Vol. 45, No. 5, 345–350

RELIABILITY, SEX DIFFERENCES, AND VALIDITY IN THE LEADERLESS GROUP DISCUSSION TECHNIQUE

WALTER A. KAESS, SAM L. WITRYOL, AND RICHARD E. NOLAN

University of Connecticut

Various technical, predictive, and theoretical characteristics of the Leaderless Group Discussion (LGD) Technique in military, industrial, and college settings have been reported recently (Bass, 1954; Borg & Tupes, 1958; Gleason, 1957). The LGD is a miniature situation in which 6 to 10 subjects discuss a general problem for approximately an hour, at the end of which they present a group solution. The subjects are usually judged by trained observers instructed to rate specified characteristics of leadership emerging in the situation, but Gleason (1957) has suggested an economical procedure in which the subjects rate each other. Our study provides a comparison of the reliability of three LGD judging techniques with analyses of the effects of sex composition of the judging, as well as of the participant, groups. Personality correlates of leadership were examined along with the judging behavior of the leaders themselves.

It would appear almost impossible to meet standard conventions of objective testing in the LGD where the problem presented to the group is intentionally ambiguous, where the personal-social composition differs from group to group, and where observer instructions vary from recording sheer talk to rating general and/or specified characteristics of leadership (Bass, 1954). Yet, rater reliability estimates have approached the limits for any typical paper-pencil test, and stability coefficients, though more variable, have been satisfactory (Bass, 1954). As research has accumulated to yield more specifications of conditions of LGD administration for objectivity, predictive efficiency, and economy, the technique has become increasingly popular among investigators. Thus, Bass concluded that groups of six to eight, whose members are strangers to each other, and who discuss case history material, yield promising results; Gleason's (1957) intersubject rating system dispenses with the necessity of trained observers. Consequences of these findings lead to logical, provocative conceptualizations of the LGD. Demonstrated validity and reliability estimates of the LGD combined with economical rating techniques constitute a good predictor instrument for leadership. The LGD remains, however, more cumbersome than simple paper-pencil tests which can be administered to large groups, but it can also be viewed as an *intermediate criterion* of leadership with better defined characteristics than widely used applied criteria. More economical scales could be validated against such a criterion before application to more expensive and complex criteria.

Specifically the purposes of the present investigation were to: (*a*) compare the degree of LGD judging agreement within and among three conditions of ranking defined as observers, recorders, and participants; (*b*) analyze the influence of sex membership of the judges and of the participants; (*c*) evaluate the accuracy of leader and nonleader rankings; and (*d*) examine the relationships between LGD leadership and selected factors on the Guilford-Zimmerman Temperament Survey.

PROCEDURE

Subjects. Subjects were students from 21 laboratory sections in introductory psychology. The number of students in most lab sections was about 30; subjects were assigned to participant and observer or recorder roles by a random method.

LGD Participants. Typically two groups of six subjects each were seated in separate circular arrangements at opposite ends of the room. Group A was all male, and Group B contained mixed sexes, three males and three females. In a few of the smaller lab sections there was only one LGD group. Participants wore identifying tags, numbered from one to six. There were 19 Group A arrangements and 19 Group B arrangements, totaling 228 LGD participants.

Observers and Recorders. Each group was observed by two pairs of nonparticipating students seated on

the periphery of the circle. Using bridge terminology, North and South were designated "observers"; East and West, "recorders"; each of the two pairs consisted of a male and a female. Recorders were provided with appropriate tabulation sheets to check every incidence of talking by each member of the LGD group. Observers were instructed to fill out a five-point rating form employed in the Bass studies (Bass & Norton, 1951) at the end of the observation period. This rating scale contains 18 items involving initiation of behavior, influencing others, consideration for others, and authoritativeness. Observers and recorders, 76 males and 76 females, were selected from their lab sections.

Task Presentation. The LGD groups listened to a short, 2-minute tape recording of a playlet concerning a coed's indecision about remaining in college. There was no obvious, pat solution. The participants in each group discussed the coed's problem 20 minutes, and then each participant wrote a solution. They next ranked each member on leadership, as did the observers and recorders, according to the following instructions:

> Place the number of the discussant who in your opinion was the leader of the group opposite Number 1; then list opposite Number 6 the number of the person who was lowest in leadership. Then opposite Number 2 list the number of the discussant next highest in leadership; opposite 5, the number of the person second lowest in leadership. Continue until all people have been ranked.

It should be noted that three independent leadership rankings were available from observers, recorders, and participants.. The experimental design provided data relevant to sex differences in judging LGD behavior, as well as in the emergence of leadership. Scores on the first five scales of the Guilford-Zimmerman Temperament Survey, administered 10 weeks prior to the LGD experiment, were available on most subjects for analysis of personality correlates.

Results

Reliability

The measures of reliability employed in this study might be classified the "degree of agreement" among judges, with analyses including the effects of the sex of the judges, the sex of the LGD participants, and the ranking condition. Two of the ranking conditions, observing and recording, were relatively passive in contrast to active LGD participation which led to a type of sociometric evaluation.

Within Conditions of Ranking. The degree of agreement was computed between observer pairs, between recorder pairs, and among participants. Spearman's rho coefficient was used to express agreement between each of the 38 observer, and between each of the 38 recorder, pairs. Kendall's coefficient of concordance, *W*, was particularly appropriate to express the degree of agreement among the six participants in each LGD group since it constitutes the mean of all possible rho correlations; thus, the unit of agreement was essentially the same for all types of observation. Table 1 presents rho and *W* coefficients in quartile values for the observer, recorder, and participant groups; the range of the median coefficients is .67–.92. If, as is usually the case, the ratings were pooled over a condition of ranking or evaluation, the reliabilities of the sums would be considerably higher; this would be particularly apparent for the participant sums which would be based on six rankings. These conservative reliability estimates, then, appear substantial, and they are comparable over conditions of ranking as demonstrated in Table 1, and as analyzed below.

Between Conditions of Ranking. After finding the high agreement within ranking conditions, it was important to determine the extent of agreement among the ranking conditions in order to obtain further clarification on which, if any, might be the most adequate measure of LGD performance. The three conditions of assessment represented by observer, recorder, and participant rankings, respectively, were compared. For each LGD group of six, the two observer rankings were pooled, the two recorder rankings were pooled, and the six participant rankings were pooled. Rho

TABLE 1

Quartile Values for rho and *W* Reliabilities

Groups Rated	Q	Raters		
		Observers rho	Recorders rho	Participants *W*
A[a]	75	.93	.97	.84
	50	.84	.89	.67
	25	.59	.65	.44
B[b]	75	.91	.98	.82
	50	.77	.92	.69
	25	.49	.79	.47

[a] $N = 6$ in each of 19 same sex Group A combinations.
[b] $N = 6$ in each of 19 mixed sex Group B combinations.

TABLE 2

MEDIAN RHO VALUES BETWEEN CONDITIONS OF RANKING

Condition Combinations	Group A (19 same sex sextets)	Group B (19 mixed sex sextets)
Observer-Recorder	.93	.86
Observer-Participant	.93	.84
Recorder-Participant	.94	.90

correlations were then computed for the 38 groups between all combinations of the three conditions; these results are summarized in the median coefficients presented in Table 2. The magnitudes of these median intercorrelations, ranging from .84 to .94, suggest that within a given LGD group the three conditions of ranking yield comparable results. These findings, moreover, verify the substantial reliability of the LGD technique as employed in our study. The base measures used in our subsequent analyses were the pooled participant composite rankings because: (*a*) they can be easily and economically obtained in an LGD situation, (*b*) they are based upon rankings by the largest number of judges, and (*c*) of the convenient tautology that the leader selected by the participants is by one definition the leader of the group.

Interactions between Sex and Conditions of Ranking. In this analysis of variance, having selected the pooled participant composite as our criterion measure, we considered: (*a*) two social conditions (19 same sex and 19 mixed sex LGD groups of six each), (*b*) two types of observation (the remaining two conditions of ranking by observers and recorders), and (*c*) sex within the conditions of observation (one male and one female observer, one male and one female recorder in each of 38 groups). Using our criterion measure we examined accuracy of observation as a function of ranking condition, sex of the observer, and interaction with the sex composition of the LGD group. The sum of the differences squared between the criterion and the rankings of an individual observer or recorder was the unit of measurement for each social condition in a double analysis of variance. The two main effects, conditions of observation and sex of the observer, did not yield significance, nor did the interactions with social composition. The findings did not support expectations based on our research in social intelligence (Witryol & Kaess, 1957) that females would be more perceptive of social behavior.

Accuracy of Leader and Nonleader Judgments

The accuracy of the leader's perception of the group as a significant requisite has been treated in a contradictory fashion in the research literature. Those investigators who deal with an empathy construct in leadership have emphasized social insight and perceptual accuracy, while other workers have found that personal characteristics like dominance and egocentricism, which may blunt accurate social perception, are more relevant features. Our data provided a modest test of these incompatible positions.

On the basis of the participant sum of scores received, those ranked highest in leadership were compared in social accuracy with those ranked lowest in each group. The criterion of group structure was defined by the rankings assigned via the pooled participant composite, the base measure indicated above. Social accuracy was then determined by the amount of agreement between an individual's judgments and the pooled composite. The sum of the differences squared between judgments by an individual of high or low leadership and the pooled composite judgments was the specific measure for expressing social accuracy of leaders and nonleaders.

From the analysis of variance in Table 3 it was inferred that nonleader judgments were more congruent with the criterion than were leader judgments at a very high statistical confidence level, beyond .001. In Group A the median rho for nonleader judgments cor-

TABLE 3

ANALYSIS OF VARIANCE BY LEADERSHIP AND LGD GROUP CONDITIONS

Conditions	*ms*	*df*	*F*
Leaders-Nonleaders	882.6	1	13.11*
LGD Groups	99.6	1	1.48
Interaction	165.6	1	2.46
Within Groups	4848.0	72	

* Significant at .001 level.

related with the pooled composite was .89; for leaders, .79; in Group B the rho's were .91 and .60, respectively ($N = 19$ in both groups). Perceptual accuracy of leadership did not yield statistical significance as a function of group conditions, nor were the interactions significant. In short, nonleaders were more accurate in perceiving LGD leadership status. Further analyses which we shall not report in detail here support this general finding. Thus, the leader was as likely to disagree with the composite in judging nonleaders, as in identifying leaders, reflecting lack of congruence with the criterion in various portions of the judging scale. Finally, leader judgments were in less agreement than nonleader ratings when both groups were compared with the observer and recorder composites.

Personality of LGD Leaders

The first five scales of the Guilford-Zimmerman Temperament Survey, administered 10 weeks before the present investigation, were available on about 70% of the subjects. In Group A (male sextets) 26 leaders, defined as Ranks 1 and 2 on the pooled composite criterion and 27 nonleaders, defined as Ranks 5 and 6, were compared on the GZ scale. Leaders were higher than nonleaders on General Activity (G), Ascendance (A), and Sociability (S) at .05, .01, and .05 confidence levels, respectively, as demonstrated in Table 4. A discriminant function equation computed for Group A took the raw score form:

$$D = .838\ \mathrm{G} + 1.858\ \mathrm{A} + .348\ \mathrm{S}$$

The F of this discriminant function is 28.22 for $df = 3/46$; the .01 confidence level, 6.60. The mean discriminant scores for leaders and nonleaders were 53.93 and 41.74, respectively. Maximum accuracy of leadership classification would follow, then, from a D score of 47.84 and above.

When the discriminant equation was applied to males in Group B (mixed sex sextets), seven of nine leaders and 12 of 17 nonleaders were correctly classified. The chi square of 5.54, significant at the .02 level, suggests that leadership personality characteristics for males in one type of LGD group (same sex) are similar in the other type (mixed sex). This is further supported by an r of .45 between D scores and the pooled composite leadership scores for 34 males in Group B.

The discriminant function did not predict female leadership in the mixed sex LGD groups, nor were scores on the five GZ scales suggestive of leadership-nonleadership differences. Personality characteristics appropriate to male leadership do not appear to be the same as those in female leaders, in mixed groups, the only ones available for the female analysis. The sex distributions over the six ranks may provide some clues. The polar Ranks 1, 5, and 6 were dominated by males: 67%, 69%, and 61%, respectively; the intermediate Ranks 2, 3, and 4 were dominated by females: 70%, 58%, and 70%, respectively. Females appear to exercise more subtle, if secondary, leadership when the sexes are mixed. Those who actively competed for the first position with males fall outside the conventional pattern. Obviously personality characteristics of females in a mixed sex situation

TABLE 4

LEADER AND NONLEADER GUILFORD-ZIMMERMAN SCORES FOR MALE GROUPS

Scale	Leaders ($N = 26$)		Nonleaders ($N = 27$)		*t*
	M	*SD*	*M*	*SD*	
General Activity	18.00	6.32	14.12	5.81	2.28*
Restraint	14.92	4.71	15.07	5.40	.11
Ascendance	17.32	4.82	13.28	5.53	2.79**
Sociability	19.36	6.25	15.20	6.95	2.25*
Emotional Stability	16.35	5.40	16.63	5.20	.17

* Significant at .05 level.
** Significant at .01 level.

are more complex in association with leadership under these circumstances.

Discussion

Our reliability findings are high and congruent with those reported in the definitive LGD review (Bass, 1954). We have also demonstrated substantial agreement among participants when they rank their own peers within a group, an analysis not previously found in the literature, and we have compared these rankings with two different external rating observations of LGD leadership. The strong agreement within and among these various conditions of observation are impressive when one considers the heterogeneous social possibilities inherent in 38 six-member groups where the members were more or less strangers to each other and where 19 of the groups contained male and female subjects. In one group a leader might take command and be subsequently dethroned, in another no one had much to say, and in a third everyone talked at the same time. These casual observations have been confirmed and systematized in a small group study employing Bale's categories (Kirscht, Lodahl, & Haire, 1959) from which the investigators concluded that the leader is the one who has ideas where they are lacking, or that the leader has ideas that are integrative and task relevant when talk abounds. Our findings strongly reflect the accuracy of LGD leadership determination despite the variability of group composition, situational demands, and the special requirements of leadership.

We are tempted to speculate about an hypothetical predisposition toward leadership which takes many forms as a function of social situations. The LGD technique yields a global approach to leadership comparable to Binet's success in the measurement of intelligence where Cattell's atomism failed. Carrying this analogy one step further, the personal correlates of small group leadership so frequently studied may be similar to the implications, with varying degrees of success, of factor analysis for intellectual measurement and theory. And in both areas the situational approach too often leads to a frustrating infinite regress. None of these approaches need be abandoned; reviews of the literature on small groups have pointed up the multiple possibilities (Bass, 1954; Mann, 1959). Global judgments of observers, however, and particularly of participants may uncover personality characteristics underlying leadership in a wider variety of social situations; flexibility of behavior required for leadership under diverse circumstances may not be discovered if emphasis is placed upon the minutiae of action.

The personality correlates of male LGD leadership identified in our investigation (general activity, ascendance, and sociability or "GAS") suggest an intriguing research definition of the issues raised. The GZ factors of ascendance and sociability for small group leadership have been cited in the two pertinent reviews by Bass and by Mann; these factors have also been confirmed with projective material (Mussen & Porter, 1959). Bass cited these correlates in female groups; we were unable to replicate for our females in mixed sex groups, but we did get replication for males in both same sex and mixed sex groupings. The evidence reflects the significance of these personality factors for males in a wide variety of LGD situations. The energetic, ascendant, sociable male is likely to be a leader among both men and women in the most global sense. The failure to replicate for females in mixed sex groupings is ascribed to cultural role expectations, but the same factors "work," according to the literature, for female leaders in female groups.

If we can assume that the elusive concept of interpersonal sensitivity was to some extent reflected in our analysis of leader and nonleader judging accuracy, and if LGD leadership represents important dimensions, the implications for research direction are fairly clear. Our leaders were inferior to the nonleaders in their accuracy of ranking both leaders and nonleaders in LGD groups, based on our sociometric and external ranking criteria. Our results support a qualified endorsement of the relevance of certain strong personality traits demonstrated in another LGD study (Borg & Tupes, 1958) where rated characteristics such as extroversion, assertiveness, and energy were related to leadership. Mann's review (1959) of 500 personality studies in-

volving small group behavior tends to support this position, and he pointed out difficulties in dealing with such constructs as empathy and interpersonal sensitivity. It may well be, on the other hand, that personality trait correlates of leadership are a function of the greater popularity of this research approach among psychologists, leading to an exaggerated emphasis as a consequence of the sheer number of investigations.

To conclude, certain requirements of economy and standardization of LGD procedures were urged by Bass (1954). Gleason (1957) reduced the need for trained observers by "buddy" ratings, and we have reduced discussion time to 20 minutes and demonstrated substantial reliability for male and mixed sex groups. Other investigators have employed brief discussion periods without reporting reliability (Borg & Tupes, 1958; Kirscht et al., 1959; Mussen & Porter, 1959). We recommend the LGD as a relatively well standardized prototype of group behavior with strong possibilities of general predictive efficiency. Assuming the latter we are impressed with the potential utility of the LGD as a stable intermediate criterion against which to assess more economical pencil-paper measures of leadership. The LGD may well represent a model paradigm for getting a criterion to stay put in a complex area of measurement.

Summary

Three Leaderless Group Discussion judging techniques involving leadership rankings by observers, recorders, and participants were applied to 228 college subjects arranged in 19 all male and 19 mixed sex groups of six each. Judging-agreement within and among ranking conditions was high, despite the fact that judges in all conditions were 380 untrained college observers; the sex of those rated and those being rated did not appear to affect ranking accuracy.

When evaluated against a pooled composite ranking criterion of participant judgments, a form of sociometric assessment, LGD leaders were inferior to nonleaders in their social accuracy of judging LGD leadership. Although leaders failed in interpersonal sensitivity as measured by the criterion above, the males were characterized by "GAS"—General Activity, Ascendance, and Sociability factor correlates from the Guilford-Zimmerman Temperament Survey; these relationships are consistent with published evidence reviewed by Bass (1954) and Mann (1959). A GAS discriminant function failed to replicate for female leaders whose LGD positions were available for analysis in mixed sex groups only, but the function was applicable to male leaders in same sex and mixed sex groups. The Bass review has indicated similar personality traits for female leaders in same sex LGD groups.

Our findings taken together with published evidence suggest a reliable and economical approach to LGD leadership assessment under complex social circumstances. The definition by personality correlates of the characteristics of LGD leadership, demonstrated in research, leads us to recommend this small group approach not only as a useful predictor, but also as a fruitful intermediate criterion for evaluating other research approaches to leadership.

References

Bass, B. M. The leaderless group discussion. *Psychol. Bull.*, 1954, **51**, 465–492.

Bass, B. M., & Norton, F. T. M. Group size and leaderless discussions. *J. appl. Psychol.*, 1951, **35**, 397–400.

Borg, W. R., & Tupes, E. C. Personality characteristics related to leadership behavior in two types of small group situational problems. *J. appl. Psychol.*, 1958, **42**, 252–256.

Gleason, W. J. Predicting army leadership ability by modified leaderless group discussion. *J. appl. Psychol.*, 1957, **41**, 231–235.

Kirscht, J. P., Lodahl, T. M., & Haire, M. Some factors in the selection of leaders by members of small groups. *J. abnorm. soc. Psychol.*, 1959, **58**, 406–408.

Mann, R. D. A review of the relationships between personality and performance in small groups. *Psychol. Bull.*, 1959, **56**, 241–270.

Mussen, P. H., & Porter, L. W. Personal motivations and self-conceptions associated with effectiveness and ineffectiveness in emergent groups. *J. abnorm. soc. Psychol.*, 1959, **59**, 23–27.

Witryol, S. L., & Kaess, W. A. Sex differences in social memory tasks. *J. abnorm. soc. Psychol.*, 1957, **54**, 343–346.

(Received December 12, 1960)

REVUE DE PSYCHOLOGIE APPLIQUÉE

PUBLICATION TRIMESTRIELLE

Directeurs : Dr P. PICHOT et P. RENNES

Cette Revue s'adresse aussi bien aux cliniciens (psychologues ou psychiatres), qu'aux psychotechniciens (orienteurs, psychologues de la profession).

Deux rubriques sont orientées vers l'application : **Techniques et Méthodes en psychologie de la profession** et **Techniques et Méthodes en psychologie clinique.** Ces rubriques ont pour but d'exposer sous une forme précise et concrète les techniques fondamentales, d'éclairer des points douteux, de présenter, même sous forme d'aide-mémoire, des méthodes pratiques de conduite des applications. Elles sont complétées par des **Revues générales** qui permettent de faire le point des recherches dans des domaines intéressant directement l'application. Dans la rubrique **Travaux originaux** prennent place des études d'ordre plus général.

Enfin les autres rubriques **Chroniques et Documentation** et **Analyses** donnent, tant sur le plan technique que sur le plan professionnel, un tableau de la vie quotidienne en psychologie appliquée

Rédaction et Administration : Square Jouvenet, PARIS (16è)

Règlements au nom de la Société André Vidal et Cie,
15, rue Henri-Heine, PARIS (16è)
C.C.P. 5851-62 PARIS

SUBSCRIPTIONS : 1 year, France : 1.500 francs - Foreign : 1.800 francs

NUMÉRO SPECIMEN SUR DEMANDE

MANAGEMENT SCIENCE

QUARTERLY JOURNAL OF THE INSTITUTE OF MANAGEMENT SCIENCES

Make TIMS your information center for new ideas in the Management Sciences

For membership and subscription information write to

THE INSTITUTE OF MANAGEMENT SCIENCES

Box 273

Pleasantville, New York

CONTEMPORARY PSYCHOLOGY

. . . APA's journal of book reviews

- Critical reviews of the most recent books in psychology and related areas
- The Editor's page: comments on news in the publishing world, on criticism, on literary style
- A letters-to-the-editor section
- Instructional Media: reviews and comments on films, TV programs, workbooks, teaching machines

AMERICAN PSYCHOLOGICAL ASSOCIATION
1333 Sixteenth St., N.W.
Washington 6, D. C.